1 OVERVIEW

This book serves as a guide for those interested in using IBM SPSS Statistics software to assist in statistical data analysis—whether as a companion to a statistics or research methods course, a stand-alone guide for a particular project, or an aid to individual learning. The images and directions used in this book come from IBM SPSS Statistics Version 29.0, first released during the summer of 2022. If you are using IBM SPSS Statistics Version 28, you will notice considerable consistency. For anyone using SPSS Versions 27, 26, 25, 24 … or even PASW Statistics 18 or earlier, there will be some consistency with these instructions and images, although there will be some areas where there are differences due to the upgrades over time in the SPSS Statistics software, including upgrades made for SPSS Version 29. So, the older your version of SPSS Statistics (or even PASW Statistics), the more likely you will see some (or more) differences in methods, availability of options, screen views, and output format.

WHAT'S THE DIFFERENCE BETWEEN SPSS STATISTICS AND PASW STATISTICS? NONE.

There is essentially no difference. The program formerly known simply as "SPSS" became "SPSS Statistics" with the release of Version 17.0 and then "PASW Statistics" with the release of Version 18.0. After SPSS, Inc., became an IBM company in October 2009, the branding was changed so that future releases of the software (Version 19.0 and beyond) would become known as "IBM SPSS Statistics." When SPSS was originally developed, it stood for "Statistical Package for the Social Sciences." The motivation for the PASW (Predictive Analytics Software) branding change was to reflect the considerable reach of the software to more business-oriented realms, although now as it stands in history, this name was used only for Version 18.

STATISTICAL SOFTWARE

The SPSS Statistics software works with several kinds of computer files: data files, output files, and syntax files. Data files are those computer files that contain the information the user intends to analyze. Output files contain the statistical analysis of these data, often displayed as tables, graphs, and/or charts. Syntax files are computer instructions that tell the SPSS Statistics software what to do. Syntax files are not used with the student version of SPSS Statistics and are dealt with as an advanced application in Chapter 12 of this book. IBM discontinued the student version of SPSS software for Version 19 (PASW was the last version for which a limited student version had been

available) but has again developed a student version starting with Version 24. The student version is nearly identical to the much more expensive standard version (and can come bundled with textbooks for a very modest amount of money). The key differences are that the student version is limited to data sets with 50 variables or fewer and 1500 cases or less.

The General Social Survey (GSS) serves as the secondary data set used throughout this book to demonstrate typical functions of the statistical software by example. IBM SPSS Statistics is the software program, produced by SPSS, an IBM company as of 2009, based in Chicago, Illinois. The GSS is a data set read and analyzed by the SPSS Statistics software; it is a data file containing the information to be analyzed. The two things are distinct and can be used in separate contexts without the other, although the GSS data file used for this book is an SPSS data file and cannot be read without opening it in SPSS Statistics or converting it to another file format suitable for use in another program.

ABOUT THE GSS DATA

The National Opinion Research Center (NORC) at the University of Chicago administers the GSS. The GSS was started in 1972 and continues today. The data used for the examples in this book come from the latest available completed version of the GSS at the time of writing, collected in 2022. According to NORC, with the exception of the U.S. Census, the GSS is the most frequently analyzed source of information in the social sciences. NORC acknowledges that the GSS has been used in more than 14,000 instances in scholarly journals, books, and doctoral dissertations. Furthermore, more than 400,000 students annually use the GSS in their work.

The GSS contains many demographic and attitudinal questions, as well as rotating topics of special interest. Several core questions have remained unchanged in each survey since 1972, which allows for rich longitudinal research about attitudes, opinions, and demographics in the United States. Topical questions appear sometimes for just 1 year; other times, they can appear for a period of years. Therefore, the GSS is versatile as a longitudinal data resource and a relevant cross-sectional resource.

To maximize the amount of information that can be collected in this massive interviewing project, the GSS uses a *split ballot design*, in which NORC asks some questions in only a random subsample of the households and asks other questions in other households. Some questions, including demographic items, were asked of all respondents. When we begin analyzing the GSS data, you will notice that some data items have substantial numbers of respondents for whom data are marked as missing. For the most part, this refers to respondents who were not asked those particular questions as a result of the split ballot design.

Although many items were asked of only subsamples of respondents, you can still take the responses as representative of the U.S. adult noninstitutionalized population, subject to normal sampling error. For more information about how the GSS data were collected, see Appendix B, "Field Work and Interviewer Specifications," and Appendix C, "General Coding Instructions," in *General Social Surveys: 1972–2018: Cumulative Codebook* (Smith et al., 2019). NORC has also started making single-year codebooks available, beginning in 2021: *2022 General Social Survey Cross-Section*

Codebook, Release 3 (Davern et al., 2024.) So, the 2022 GSS release used throughout this book has an associated single-year codebook that you can use. This takes away the longitudinal frequency data available in the cumulative codebook but provides a slightly more user-friendly approach.

SPSS/PASW ELECTRONIC FILES

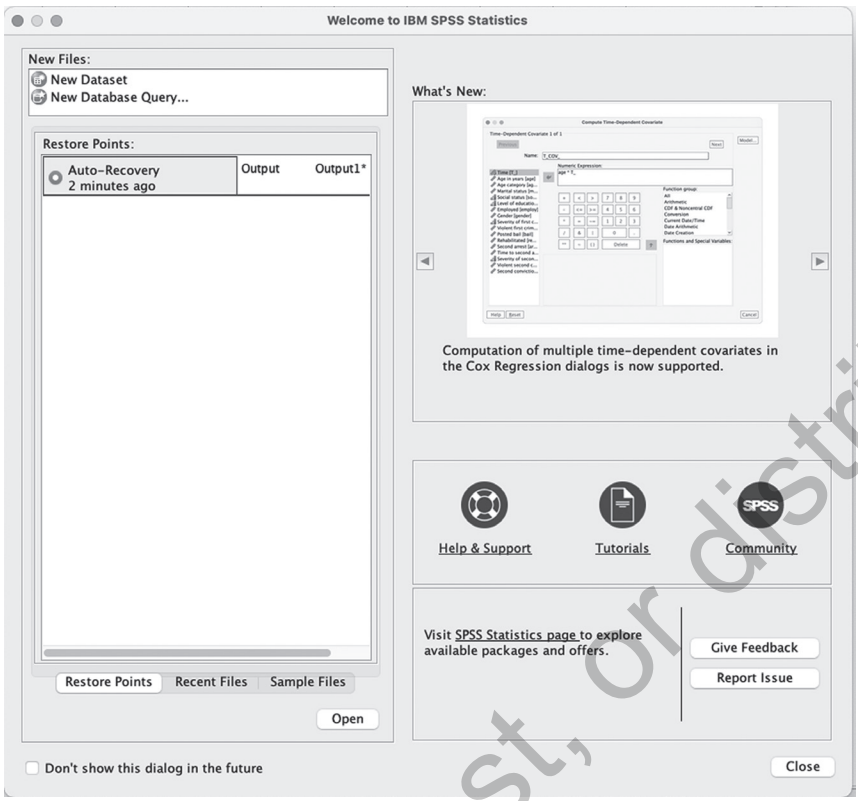
IBM SPSS Statistics 29 (as well as other versions of SPSS/PASW Statistics) uses different file extensions, or endings, and associated icons to signify types of files. For instance, a file named "file.sav" is a data file called "file." The.sav extension is used to signify that it is a data file. Again, data files contain the information SPSS/PASW Statistics analyzes. SPSS can also open files from another statistical software program (Stata) automatically, without the need for importing and specifying details; those files are labeled with a "*.dta" extension, which also signifies a data file. So, "file.dta" is also a data file called file created in or for Stata. Again, SPSS will automatically open those files, including the data and properties. However, to save the file after you've worked with it in SPSS (assuming you've made any changes at all), you'll need to save it with a *.sav file extension.

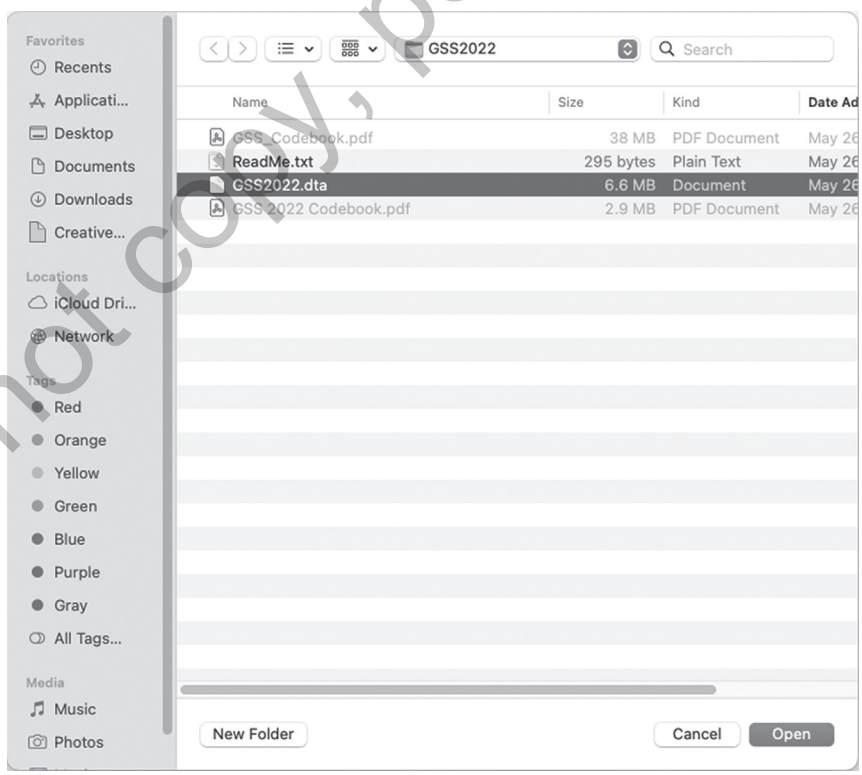
A file with the extension.sps is an SPSS Statistics syntax file, and a file with the extension.spv (or.spo for versions of SPSS software prior to Version 17.0) is an SPSS Statistics output file. Output files contain analyses of data, such as charts, tables, and other statistical and data manipulation information. Syntax files contain coded instructions for SPSS Statistics to perform operations on data and produce output. It is not necessary to create, save, or even deal with syntax files for most basic SPSS Statistics functions; therefore, syntax files will be covered only to the level of description and simple use in Chapter 12.

OPENING EXISTING DATA FILES

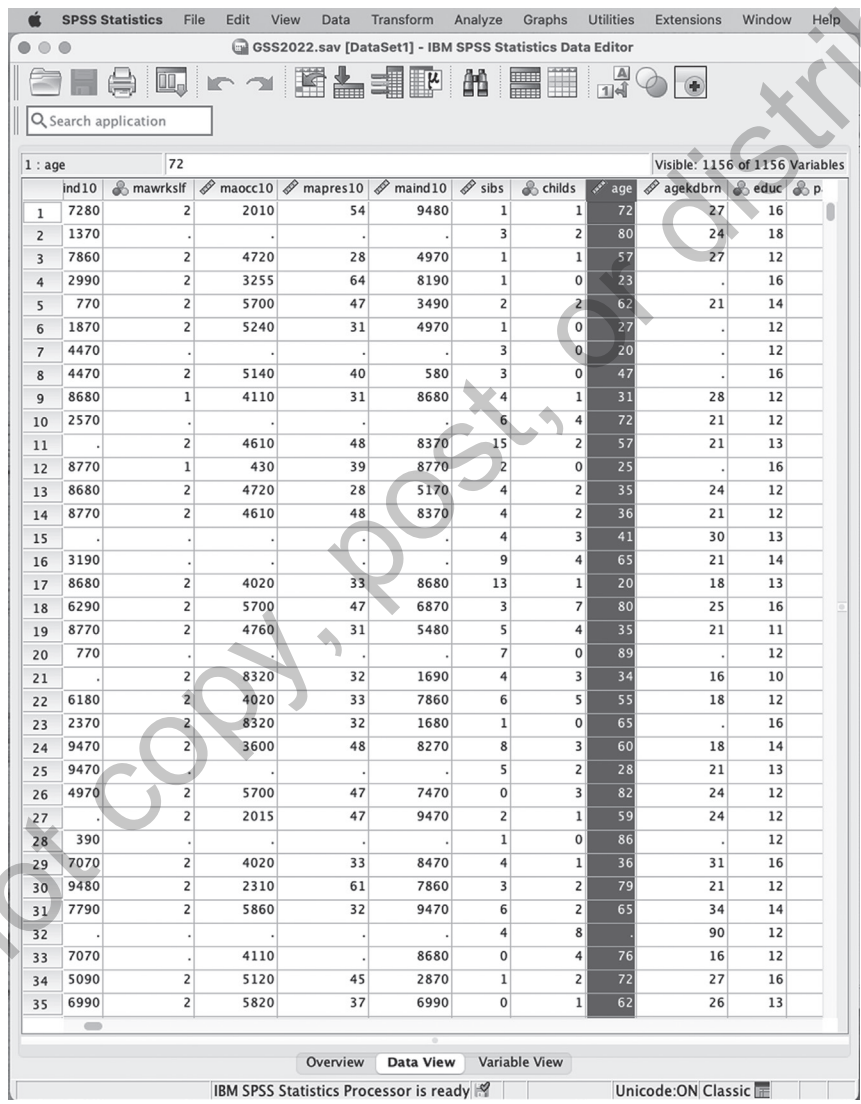
When first opening the SPSS Statistics 29 application, you will be presented with a window, like the one in the first screen image below, that offers options for how to begin your SPSS session. On the lower left side, you'll see a box labeled "Recent Files." You will be able to instantly click and begin working with one of those files. If you do not see the file you need, you can click the folder named "Open another file . . ." at the bottom of the list, which will allow you to locate your file. Sometimes, however, this window doesn't appear; you can see that there's a check box that, when checked, will prevent it from appearing when SPSS Statistics is reopened. (To "uncheck" the box and make it so the window will open each time the SPSS Statistics application is launched, visit the "Options" menu, described later in this chapter.)

Otherwise, to open an SPSS Statistics data file that you already have or have obtained, select the "File" menu; then choose "Open" and select "Data . . ." (For other file types, see the section on importing data from non-SPSS file formats.) At this point, you will need to navigate the disk drives attached to your computer (or network drives or other sorts of storage devices) to locate the data file you wish to open. Once you locate the file, either double-click it or click it once and then click the "Open" button toward the bottom right side of the "Open Data" dialog box.





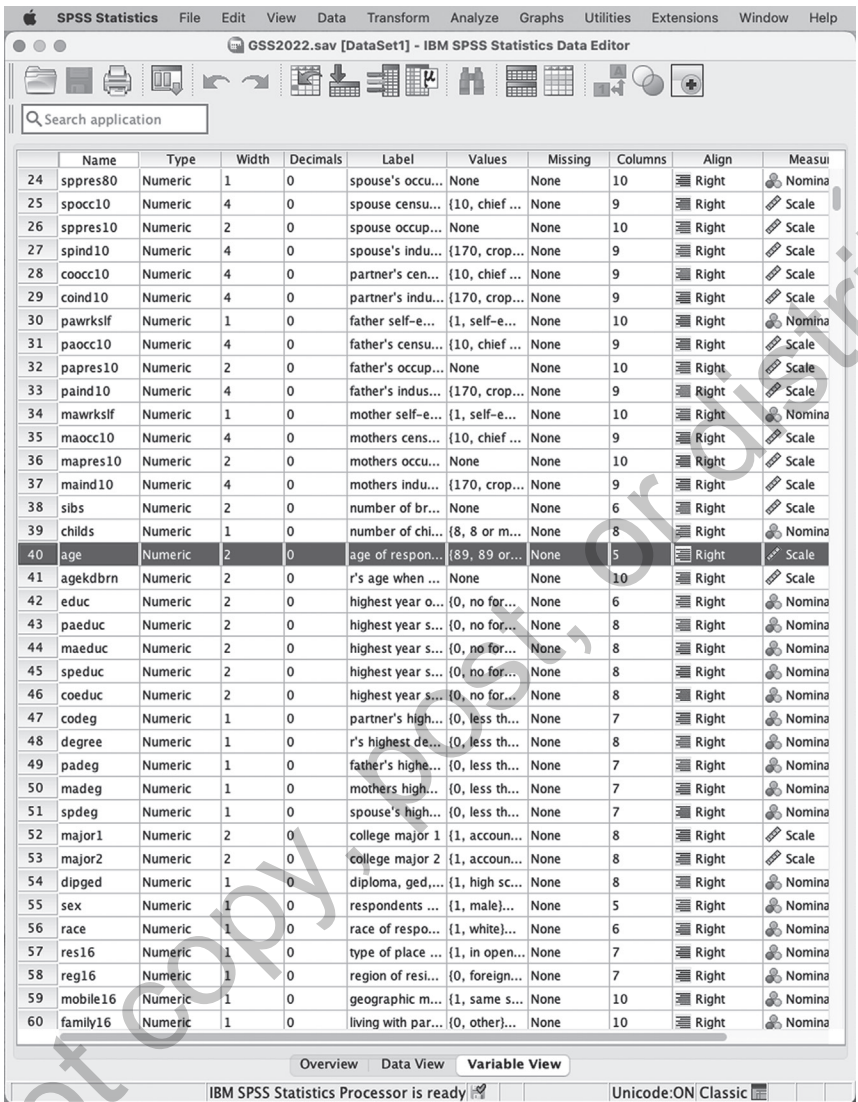
SPSS Statistics will then open the data file, and you will be presented with the information in a grid format (somewhat similar to a Microsoft Excel or other spreadsheet environment). You have choices about both the way the information is presented and the information you see. For example, you can choose to see the Data View, presented in the following screen image. Here, you are viewing the actual data. Note that the variables are listed in columns, with each case recorded as a row. The variable "AGE" has been selected as a reference point. The data in that column reveal the age of each respondent.



	ind10	mawrkslf	maocc10	mapres10	maind10	sibs	childs	age	agekdbrn	educ	p
1	7280	2	2010	54	9480	1	1	72	27	16	
2	1370	.	.	.	.	3	2	80	24	18	
3	7860	2	4720	28	4970	1	1	57	27	12	
4	2990	2	3255	64	8190	1	0	23	.	16	
5	770	2	5700	47	3490	2	2	62	21	14	
6	1870	2	5240	31	4970	1	0	27	.	12	
7	4470	.	.	.	.	3	0	20	.	12	
8	4470	2	5140	40	580	3	0	47	.	16	
9	8680	1	4110	31	8680	4	1	31	28	12	
10	2570	.	.	.	.	6	4	72	21	12	
11	.	2	4610	48	8370	15	2	57	21	13	
12	8770	1	430	39	8770	2	0	25	.	16	
13	8680	2	4720	28	5170	4	2	35	24	12	
14	8770	2	4610	48	8370	4	2	36	21	12	
15	.	.	.	.	.	4	3	41	30	13	
16	3190	.	.	.	.	9	4	65	21	14	
17	8680	2	4020	33	8680	13	1	20	18	13	
18	6290	2	5700	47	6870	3	7	80	25	16	
19	8770	2	4760	31	5480	5	4	35	21	11	
20	770	.	.	.	.	7	0	89	.	12	
21	.	2	8320	32	1690	4	3	34	16	10	
22	6180	2	4020	33	7860	6	5	55	18	12	
23	2370	2	8320	32	1680	1	0	65	.	16	
24	9470	2	3600	48	8270	8	3	60	18	14	
25	9470	.	.	.	.	5	2	28	21	13	
26	4970	2	5700	47	7470	0	3	82	24	12	
27	.	2	2015	47	9470	2	1	59	24	12	
28	390	.	.	.	.	1	0	86	.	12	
29	7070	2	4020	33	8470	4	1	36	31	16	
30	9480	2	2310	61	7860	3	2	79	21	12	
31	7790	2	5860	32	9470	6	2	65	34	14	
32	.	.	.	.	.	4	8	.	90	12	
33	7070	.	4110	.	8680	0	4	76	16	12	
34	5090	2	5120	45	2870	1	2	72	27	16	
35	6990	2	5820	37	6990	0	1	62	26	13	

Now, click the "Variable View" tab, which is located toward the bottom left of the screen or toward the bottom middle of the screen depending on whether you are using a computer with the Mac OS or Windows operating system. (Note that the "Data View" tab is currently selected; this is the default when opening a new file with SPSS Statistics 17 or

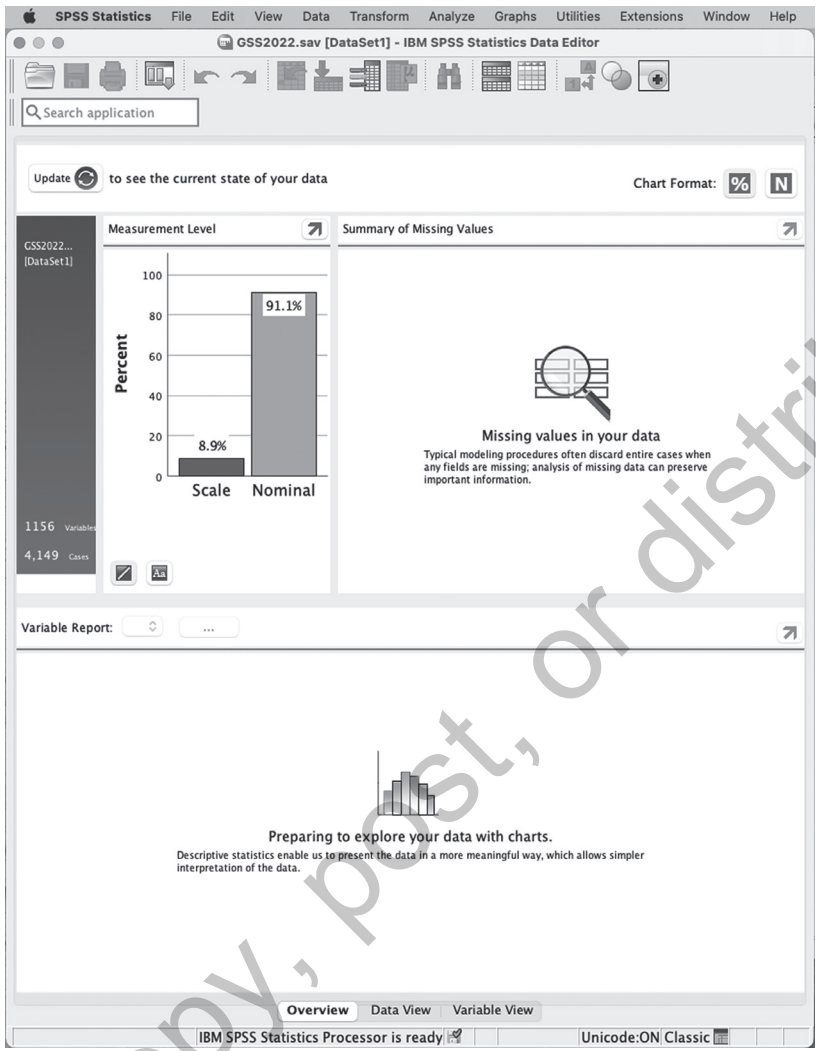
newer.) Although the information looks somewhat different, you are still looking at the same data file. See the following screen image:



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measu
24	sppres80	Numeric	1	0	spouse's occu...	None	None	10	Right	Nomina
25	spocc10	Numeric	4	0	spouse censu...	{10, chief ...	None	9	Right	Scale
26	sppres10	Numeric	2	0	spouse occup...	None	None	10	Right	Scale
27	spind10	Numeric	4	0	spouse's indu...	{170, crop...	None	9	Right	Scale
28	coocc10	Numeric	4	0	partner's cen...	{10, chief ...	None	9	Right	Scale
29	coind10	Numeric	4	0	partner's indu...	{170, crop...	None	9	Right	Scale
30	pawrkslf	Numeric	1	0	father self-e...	{1, self-e...	None	10	Right	Nomina
31	paocc10	Numeric	4	0	father's censu...	{10, chief ...	None	9	Right	Scale
32	papres10	Numeric	2	0	father's occup...	None	None	10	Right	Scale
33	paind10	Numeric	4	0	father's indus...	{170, crop...	None	9	Right	Scale
34	mawrkslf	Numeric	1	0	mother self-e...	{1, self-e...	None	10	Right	Nomina
35	maocc10	Numeric	4	0	mothers cens...	{10, chief ...	None	9	Right	Scale
36	mapres10	Numeric	2	0	mothers occu...	None	None	10	Right	Scale
37	maind10	Numeric	4	0	mothers indu...	{170, crop...	None	9	Right	Scale
38	sibs	Numeric	2	0	number of br...	None	None	6	Right	Scale
39	childs	Numeric	1	0	number of chi...	{8, 8 or m...	None	8	Right	Nomina
40	age	Numeric	2	0	age of respon...	{89, 89 or...	None	5	Right	Scale
41	agekdbrn	Numeric	2	0	r's age when ...	None	None	10	Right	Scale
42	educ	Numeric	2	0	highest year o...	{0, no for...	None	6	Right	Nomina
43	paeduc	Numeric	2	0	highest year s...	{0, no for...	None	8	Right	Nomina
44	maeduc	Numeric	2	0	highest year s...	{0, no for...	None	8	Right	Nomina
45	speduc	Numeric	2	0	highest year s...	{0, no for...	None	8	Right	Nomina
46	coeduc	Numeric	2	0	highest year s...	{0, no for...	None	8	Right	Nomina
47	codeg	Numeric	1	0	partner's high...	{0, less th...	None	7	Right	Nomina
48	degree	Numeric	1	0	r's highest de...	{0, less th...	None	8	Right	Nomina
49	padeg	Numeric	1	0	father's highe...	{0, less th...	None	7	Right	Nomina
50	madeg	Numeric	1	0	mothers high...	{0, less th...	None	7	Right	Nomina
51	spdeg	Numeric	1	0	spouse's high...	{0, less th...	None	7	Right	Nomina
52	major1	Numeric	2	0	college major 1	{1, accoun...	None	8	Right	Scale
53	major2	Numeric	2	0	college major 2	{1, accoun...	None	8	Right	Scale
54	dipged	Numeric	1	0	diploma, ged,...	{1, high sc...	None	8	Right	Nomina
55	sex	Numeric	1	0	respondents ...	{1, male}...	None	5	Right	Nomina
56	race	Numeric	1	0	race of respo...	{1, white}...	None	6	Right	Nomina
57	res16	Numeric	1	0	type of place ...	{1, in open...	None	7	Right	Nomina
58	reg16	Numeric	1	0	region of resi...	{0, foreign...	None	7	Right	Nomina
59	mobile16	Numeric	1	0	geographic m...	{1, same s...	None	10	Right	Nomina
60	family16	Numeric	1	0	living with par...	{0, other}...	None	10	Right	Nomina

Again, the "AGE" variable has been selected for reference. In this view, variables are depicted in rows, with each row showing information about a single variable, such as variable label, category label, type, level of measurement, and so forth. You can add to, edit, or delete any of the variable information contained in this view by directly typing into the cells. This view does not show the actual response data; to view those data, you would need to select the "Data View" tab.

A newer feature in SPSS Statistics is a third choice among those tabs at the bottom of the data screen: "Overview". By default, you will be provided with the number of variables and number of cases, as well as with a percentage graph showing the distribution of variables by level of measurement.

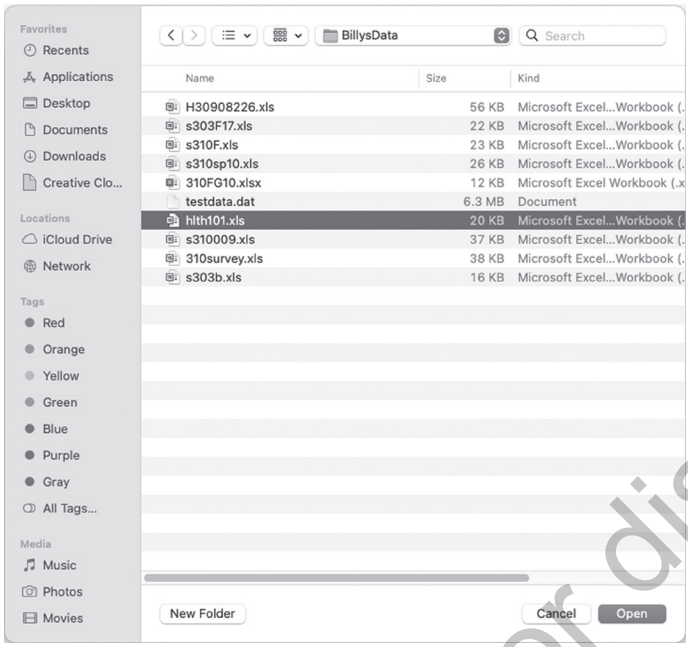


IMPORTING DATA FROM STATISTICS FILE FORMATS OTHER THAN SPSS OR PASW

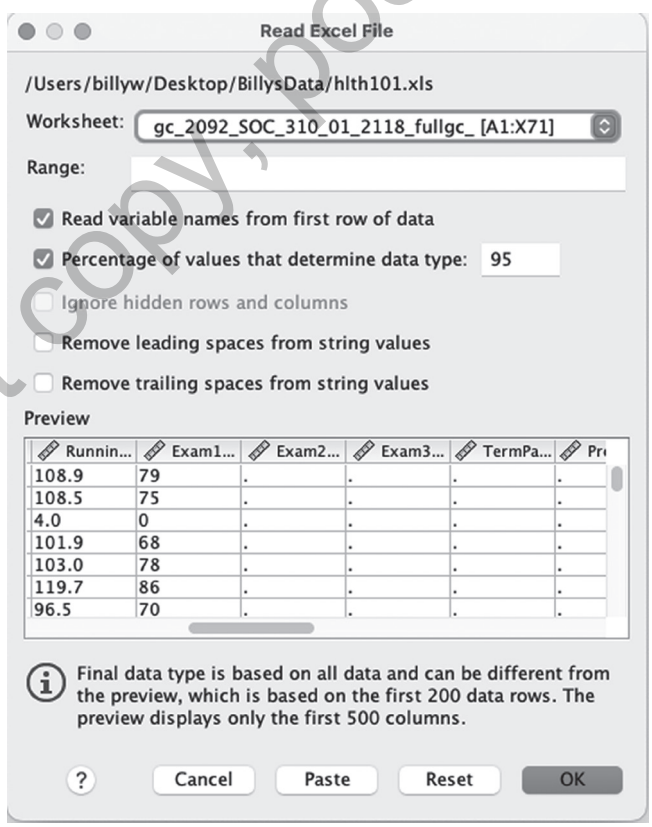
There is often a need to analyze existing data files that were not created or formatted by SPSS Statistics software. These files might be created by other statistical software packages (e.g., SAS or Stata) or by other types of numeric operational programs (e.g., Microsoft Excel). As noted earlier in this chapter, *.dta files (Stata) can be imported automatically by opening the file as if it were and SPSS *.sav file. To open an Excel file, for example, first select the same menu options you would if you were opening an SPSS/PASW Statistics data file:

File → Open → Data . . .

Now, at the "Files of type" prompt at the bottom of the dialog box, click the arrow at the right to expand the choices. Next, select "Excel (*.xls, *.xlsx, *.xlsm)." You will need to



navigate your hard drive, other drives, or other locations to find your file. Once you locate the file, select and open it. At this point, you will be presented with a new dialog box:



If the column headings in the Excel file contain the variable names, make sure the box labeled "Read variable names from first row of data" is selected. If the column headings are not formatted in a way that conforms to the SPSS Statistics variable-naming conventions, they will be transformed into permitted variable names, and the original names will be recorded as variable labels.

To import only a portion of the Excel file, enter the range of cells from which you would like to import data.

It is also possible to import data from databases, text files, and other sources. Follow the same instructions as for Excel files, except for the file type you select. Depending on the file type you choose, you will be presented with different dialog boxes or wizards to import the data and process them for use by SPSS Statistics.

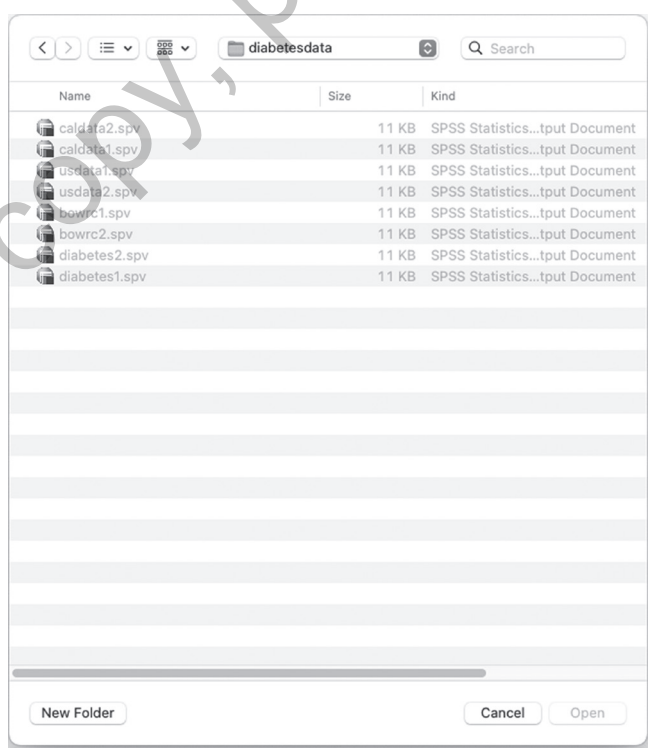
In some cases, you may simply have unlabeled data in a particular file, or the variable names or other information may be of little or no use to you. In such a case, depending on the size of the file, you could copy the data from the original numeric program (e.g., Microsoft Excel) and then paste them directly into the SPSS Statistics Data View window. This is particularly useful if you just want to add numeric values from another source and enter or program the other information using SPSS Statistics.

OPENING PREVIOUSLY CREATED OUTPUT FILES

To open a previously created output file, select from the "File" menu as follows:

File → Open → Output . . .

You will be presented with the following dialog box:



Here, just navigate to find your file, as you would any other type of file. Once you locate it, select and open the file. SPSS Statistics will open the file into an output Viewer window. There, you can view and edit it. Note that IBM SPSS Statistics Version 25 (as well as prior versions going back as far as SPSS Statistics 17) uses the.spv file extension. Versions prior to 17 of SPSS software use the.spo file extension. IBM SPSS Statistics Version 25 is capable of opening output files created with older versions of SPSS software, but files created in the new version cannot necessarily be opened in the older versions.

SAVING FILES

All types of SPSS Statistics files are saved in the same way as files in any modern computer program. Select either

File → Save to save the file with the currently assigned name

or

File → Save As . . . to save the file in a different file, with a new name.

The first option will automatically save the file without prompting you with a dialog box, unless you are working with a new, yet unnamed file. In that case, you will get the same type of dialog box as though you had selected the "Save As . . ." option. If you do choose the second option, you will be shown a dialog box prompting you to name the file and to select the location on your computer or network where the file is to be placed.

CREATING NEW SPSS STATISTICS DATA FILES

To create a new SPSS Statistics data file, select the "File" menu; then choose "New," and select "Data":

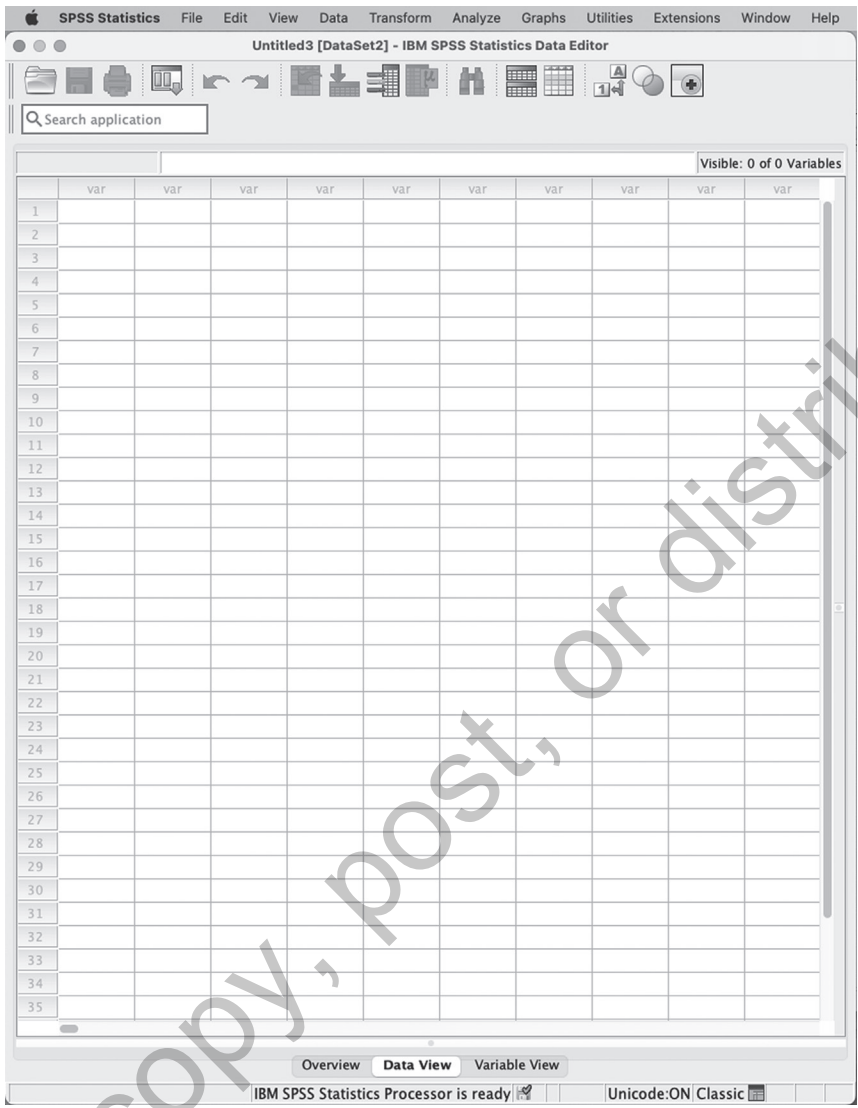
File → New → Data

You will then be presented with a blank Data Editor window like the one in the first screen image below.

You can immediately start entering information related to the variables you wish to create and/or the actual data codes you may have. In the Data Editor window shown in the second screen image, the Variable View tab has been selected, and information has been entered for two variables: "age" and "sex." Notice that the labels have been entered: "Age of Respondent" and "Sex of Respondent," respectively. Other information about the variables has been selected and entered as well.

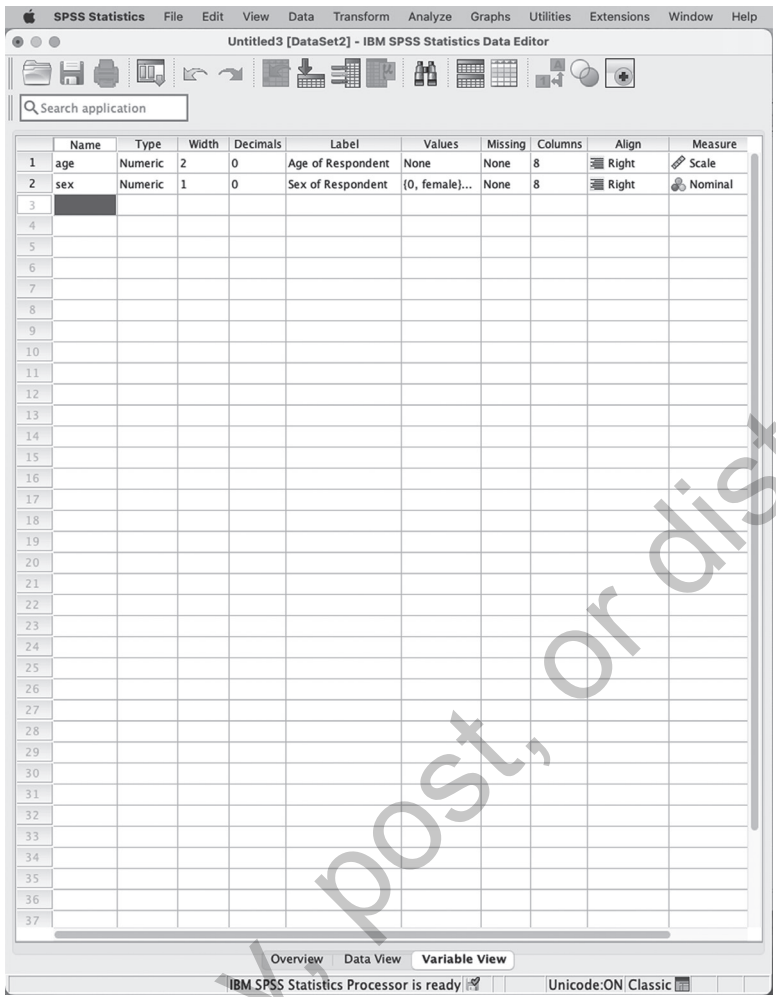
For the "sex" variable, value labels have been entered. This was done by clicking the "Values" cell for that variable and then selecting the button with three small dots. The dialog box shown at the bottom of the third screen image appeared.

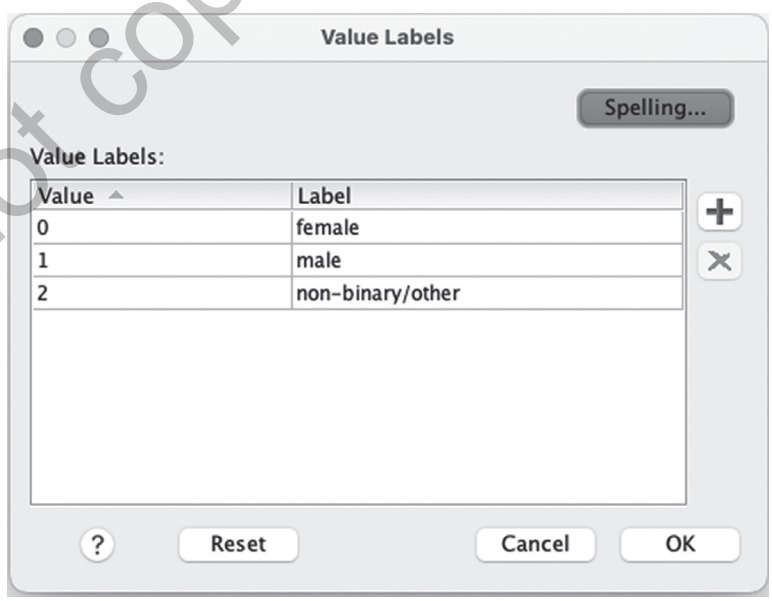
In the "Value Labels" box, you can enter the label for each category code for the variable. In this case, after clicking the "+" button on the right side of the window, "0" was entered in the "Value" box, and "female" was entered in the "Label" box. Notice



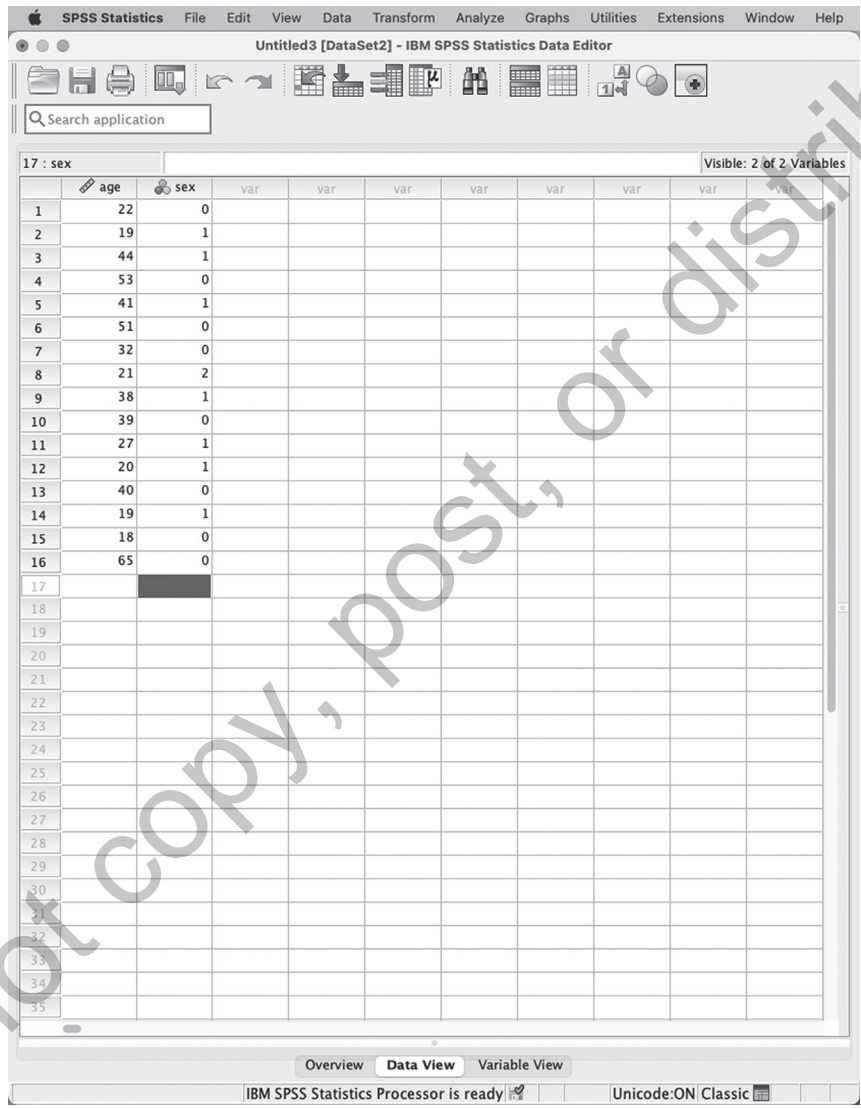
also that "1" was entered in the "Value" box, and "male" was entered in the "Label" box, after again clicking the "+" button on the right side of this window. Finally, after one more click of the "+" button, 3 was added for "Value" and "non-binary/other" was added for "Label". By doing this, we are assigning numbers to the categories of the variable "sex" so that SPSS Statistics knows how to record whether each respondent is male, female, or non-binary/other; the SPSS software uses numbers to keep track of those attributes. By using numbers, the software can track the categories and use that information to perform statistical operations such as those described throughout this book.

Data can also be entered directly into the Data Editor (Data View mode). To do this, click the Data View tab from among the three tabs (Overiew, Data View, and Variable View) at the bottom of the Data Editor window. The columns now



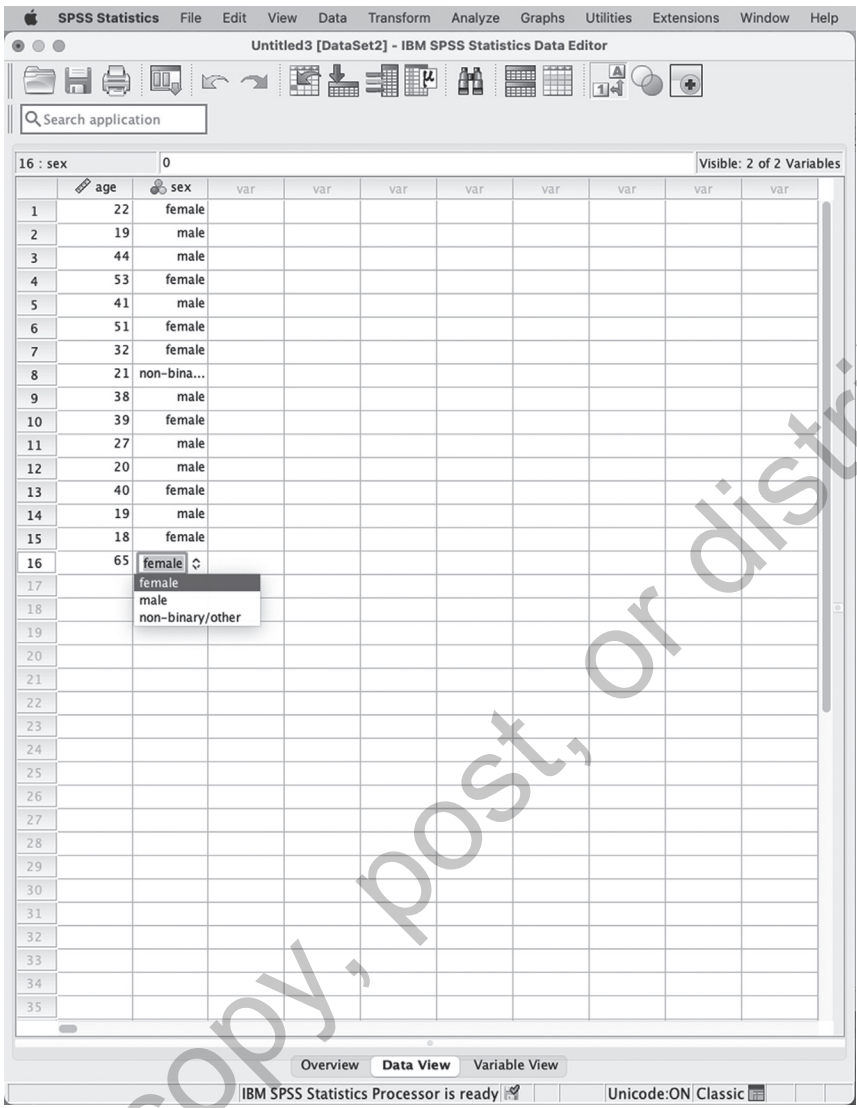


represent the newly created variables "age" and "sex." You can directly type the values for each case or row. You could also copy (<Ctrl> + C on a Microsoft Windows PC, or <Apple> + C on a Macintosh computer) and paste (<Ctrl> + V on a Windows PC, or <Apple> + V on a Macintosh computer) these values from another software program, such as Microsoft Excel, if you already have them categorized by the same variables.



Notice that the data entered (or pasted) above appear as the numeric codes assigned for nominal and ordinal variables that have those assignments. There is also a way to display the actual label in this window instead of the numeric codes. Click the following from the SPSS menus, and note the differences in the following screen image:

View → Value Labels



Value labels can also be displayed prior to entering data. When you double-click the right side of a cell, an arrow will appear. If you click the arrow, as seen in the screen image above, a pull-down menu consisting of all of the available categories for that variable will appear: In this case, the list consists of "female" and "male." If you are entering data directly into SPSS Statistics, using this option can make data entry easier and can help avoid errors, such as typos of values that are not within the range of categories for a variable.

CREATING AND EDITING SPSS STATISTICS OUTPUT FILES

Output files are created by SPSS Statistics when you instruct the software to perform functions. For example, if you request SPSS Statistics to provide frequencies and central tendency values for three variables from your data set, an output file will be produced automatically. The information you have requested will be presented in the output

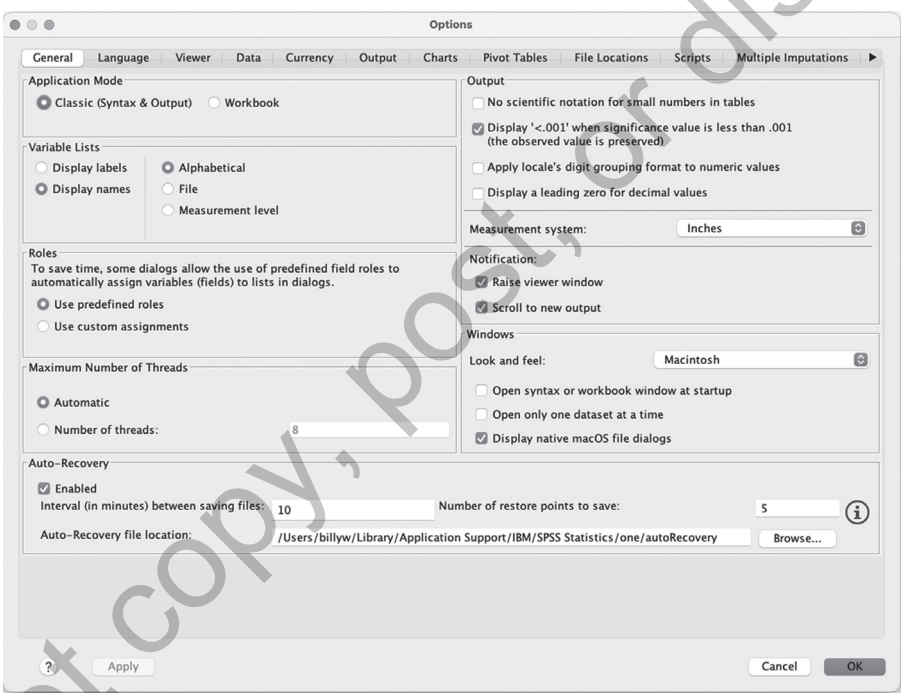
Viewer window unless an output file is already open in the output Viewer; in which case, the new information will be appended to that file. To edit the output, you'll select and double-click the part you wish to work with, and there are tools to facilitate that task. More information on this topic is provided in Chapter 4, "Organization and Presentation of Information."

PREFERENCES: GETTING STARTED

To change the settings, parameters, and preferences for the SPSS Statistics program, select the "Edit" menu and choose "Options . . .":

Edit → Options . . .

You will be presented with a dialog box like the one shown here:

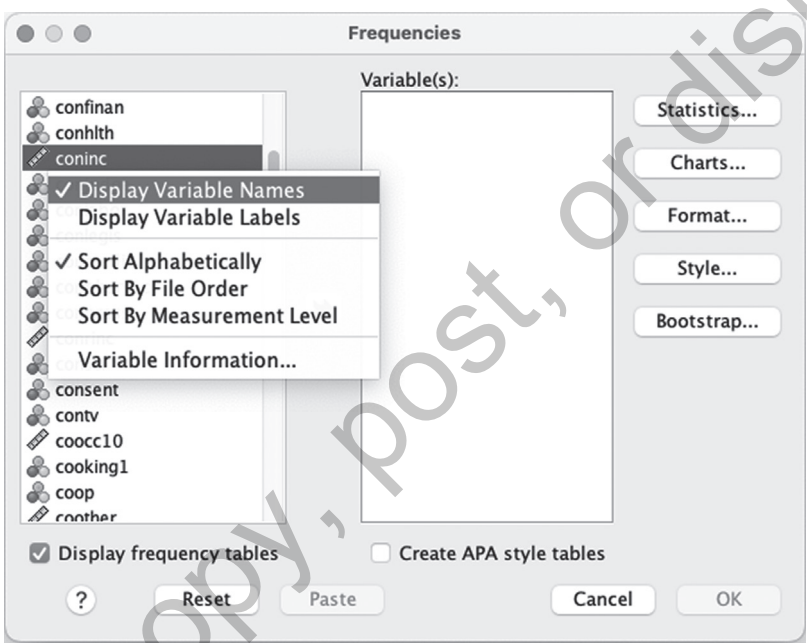


Numerous features can be controlled using this dialog box, and most are intuitive in their operation. As a user becomes more experienced, they often use more of these features. From the start, however, most SPSS Statistics users will want to make sure that variables will be displayed throughout the program in alphabetical order and by name (rather than by label). This can be done by selecting the "General" tab and clicking the radio buttons for "Display names" and "Alphabetical."

This process is particularly important if you are using or creating a data set that contains many variables, such as the GSS. Although alphabetizing the list will clearly facilitate easier access to variables, listing by name is also crucial because variable labels are more detailed and may not necessarily begin with or even use the same letters as the

variable names. Changing or verifying these settings up front can save a good deal of time and frustration. If a data set is opened and the preferences have not been set to the desired parameters, you can still go to the "Options" dialog box and make the change while the data set is open. (In some older versions of SPSS, it would be necessary to close the data set, make the change, and reopen the data set.)

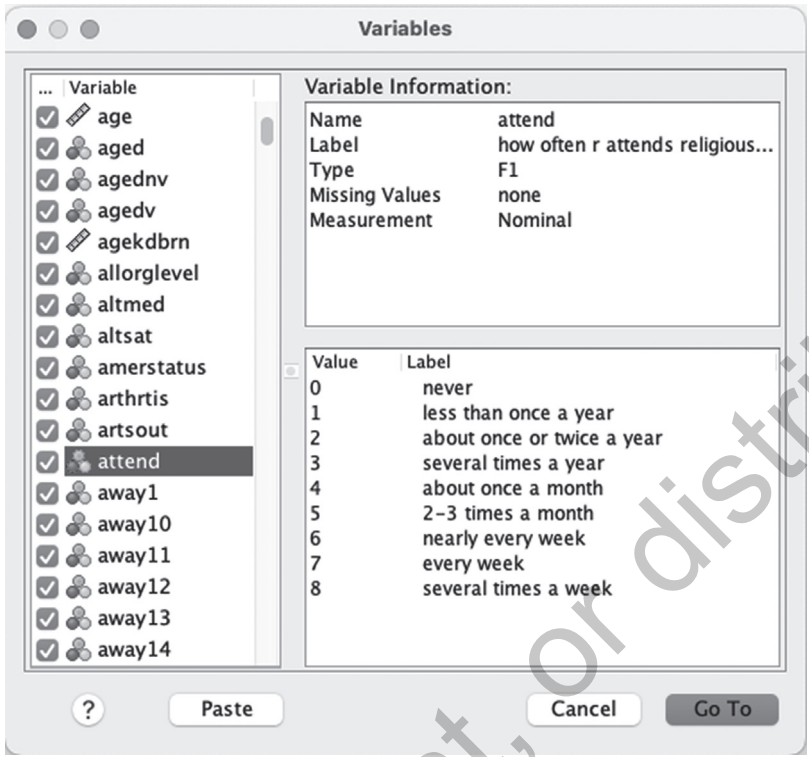
If you have started a procedure and don't wish to change the program preferences as described earlier, you can do it on the fly. Just right-click (or <Control> + click for a Macintosh computer with a standard one-button mouse) one or more of the variables from the list shown in the dialog box with which you are working. You can select "Display Variable Names" and "Sort Alphabetically." Incidentally, you can also select "Variable Information . . ." for any of the variables (not more than one at a time), which will provide details about values, value labels, and so forth.



Note that this procedure, pictured above, will make the change only in the current dialog box (including return visits to that same dialog box, "Frequencies" in this case). You will need to use the same procedure to organize variables in other dialog boxes or make the change in the program preferences window (Edit → Options). Also, note that this procedure does not work in all windows (e.g., Utilities → Variables; see the next paragraph).

To get a quick overview of the variables in a given data set, SPSS Statistics has a variable utility window to provide useful information about each variable in a way that can be easily navigated (and such that information can be easily pasted to output if desired). Opening this window will demonstrate the importance of organized naming and ordering of variables in a large data file. Choose the "Utilities" menu; then select "Variables . . .":

Utilities → Variables . . .



When you select a variable from the alphabetized list of variable names on the left, information about that particular variable will appear on the right side of the box, including the label, the level of measurement, and the value labels. This is a fast way to determine what kinds of variables are available in your data set that are suited to different statistical methods of analysis.

MEASUREMENT OF VARIABLES USING SPSS STATISTICS

Whether creating a new data file with SPSS Statistics or using an existing data file, it is important to understand how variables have been measured, or "treated," by the creator of the data file. This treatment is a factor of how the data were collected—how much information is contained within the data set about each variable.

First, it is important to be aware that SPSS Statistics can record variables as either *string* variables or *numeric* variables. String variables can consist of letters and/or numbers and cannot be treated numerically; therefore, string variables must be treated at the nominal level of measurement. Numeric variables use numbers to represent response values. These numbers may represent actual numbers, ranked categories, or unranked categories. In other words, numeric variables may be *nominal, ordinal, interval*, or *ratio*.

In social science statistics and research methods courses, variables are typically described using these four categories. Many textbooks, such as *Investigating the Social World* (Schutt, 2023) and *Adventures in Social Research* (Babbie, Wagner, & Zaino, 2022), elaborate all four of those categories. In some texts, interval and ratio measures are combined, as is the case in *Social Statistics for a Diverse Society* (Frankfort-Nachmias, Leon-Guerrero & Davis, 2026).

SPSS Statistics uses the following codes for levels of measurement: nominal, ordinal, and scale. You can select the level of measurement from the pull-down menu for each variable in the "Measure" column of the Variable View window. "Nominal" and "Ordinal" both correspond to the concepts with the same names. The "Scale" denotation corresponds to interval-ratio, interval, and ratio. These functions within the SPSS Statistics software will limit your ability to conduct analyses or create graphs on the basis of the recognized level of measurement. Therefore, it is crucial to verify that the indicator in the "Measure" column of the Variable View is correct for all variables you will use in your analyses.

2 TRANSFORMING VARIABLES

In this chapter, tools for restructuring variables are introduced. IBM SPSS Statistics allows numerous ways to reconfigure, combine, and compute variables in a data set.

RECODING AND COMPUTING VARIABLES

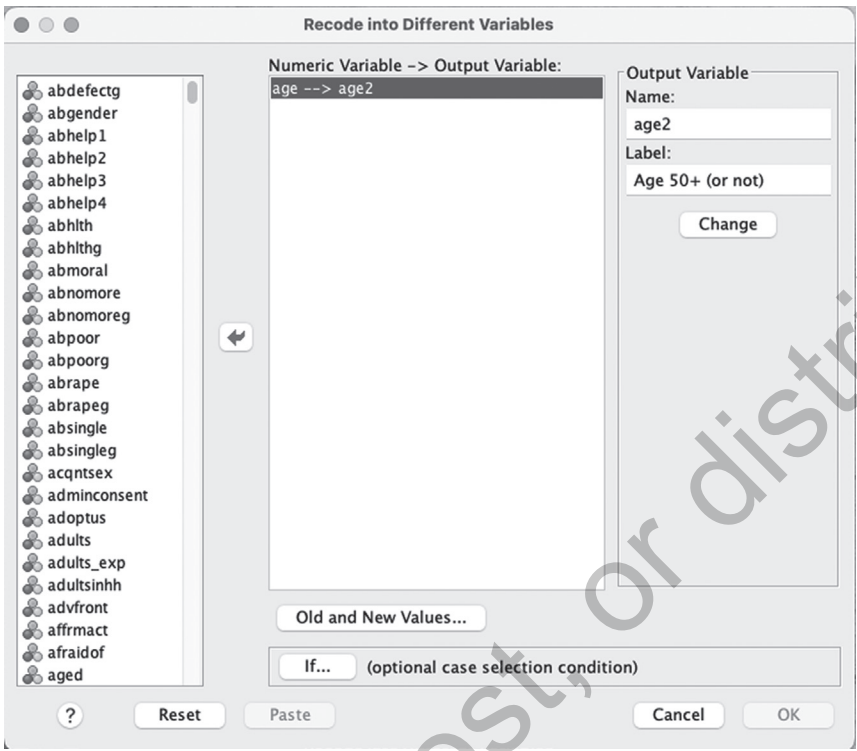
Often, one must reorganize the way data are recorded before performing statistical analyses. This might be due to the level of measurement of a particular variable a researcher wishes to change, or it could be related to the researcher's intended use of a variable. One may wish to collapse a few categories of a variable into one for appropriateness of analysis. For example, within the marital status variable, one might combine the "married" and "separated" categories to form "legally married" and combine "divorced" with "single, never married" to form "unmarried." "Recode" is the SPSS Statistics function that allows a researcher to recategorize a variable to suit the needs of the analysis.

There are many times when a researcher needs to produce a new variable from existing information in a data set but that information is not contained solely within one variable. SPSS Statistics has a "Compute" function that allows a user to both perform mathematical operations on variable data and combine data from multiple distinct variables within the file.

For this recoding example, we will use the General Social Survey (GSS) 2022 data set. We will take a straightforward case of dichotomizing age from a ratio variable, presenting the respondent's actual age at the time of interview, into just two categories with a cut point of 50 years of age. This would allow the researcher to present ages in a small frequency table, whereas before recoding age, there are so many categories (18 to 89 years of age) that presentation in a frequency table is not feasible (though, of course, there are graphics that could be used to illustrate age without recoding; see Chapter 5, "Charts and Graphs"). To recode, or change, the categories of a variable, select the "Transform" menu, and then choose the "Recode into Different Variables . . ." option:

Transform → Recode into Different Variables . . .

You will then be shown a dialog box like the one displayed in the following image:



It is useful and proactive against data loss to recode into "different variables" rather than "same variables" if you are reducing information contained within the variable. For instance, if you are recoding a ratio-level variable into an ordinal or a dichotomous variable, you would want to create a different variable. The reason is that the lost information resulting from the recoding would still be retained in the original variable should you want to change the way in which you recode the variable at some later point in time or should you determine that you need the more detailed ratio-level information for your analysis.

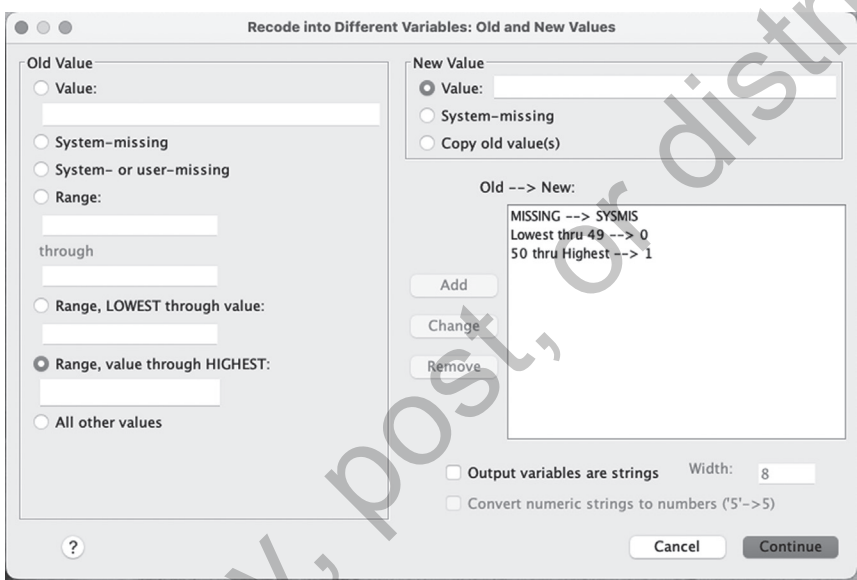
RECODING VARIABLES: DICHOTOMIES AND DUMMY VARIABLES

In this example, we recode the age variable into a dichotomy: a variable with exactly two categories, not including missing values. A dummy variable is a dichotomy usually coded with a value of "1" to indicate the existence of a particular attribute. All other attributes are coded "0" as an indication that the particular attribute represented by "1," usually the name of the variable itself, is not present. Dummy variables are particularly useful for including nominal-level variables in analyses requiring interval- or ratio-level variables to perform statistical operations, including multiple regression models (see Chapter 8, "Correlation and Regression Analysis"). A series of dummy variables can also be used to represent several attributes from a given nominal variable.

To dichotomize the age variable, first select "age" from the list of variables in the data set on the left, and then click the arrow to drag it to the "Numeric Variable" pane. Now, create a new name for the variable; in this example, the mundane "age2" has been

conjured. You may also select a label at this time, or you can attend to that at a later time through the Variable View tab of the Data Editor screen. For this example, the label "Age 50+ (or not)" was added. The next step is to click the "Change" button. This will enter the name "age2" into the "Output Variable" location, as has been done in the screen image in the second screen image in this section. Until the "Change" button is clicked, a question mark will act as a placeholder.

Now, it is necessary to give SPSS Statistics instructions for *how* the variable is to be recoded. In this example, we want to change all ages up to and including 49 into a category called "0," and all ages 50 and greater into another category called "1." Click the "Old and New Values . . ." button in the "Recode into Different Variables" dialog box, and another dialog box will appear on top, like the one that follows:



To implement the changes, first, under "Old Value," select the radio button that reads "Range, LOWEST through value." Then enter "49" in the box underneath. Then, under "New Value," select "Value" and enter "0." Now, click the "Add" button. This instructs SPSS Statistics to transform all ages up to and including 49 into the category "0."

Next, under "Old Value," select the radio button associated with "Range, value through HIGHEST." Enter "50" in the box beneath that heading. Then, under "New Value," select "Value" and enter "1." Again, click the "Add" button. This now instructs SPSS Statistics to transform all ages 50 and beyond into the category "1."

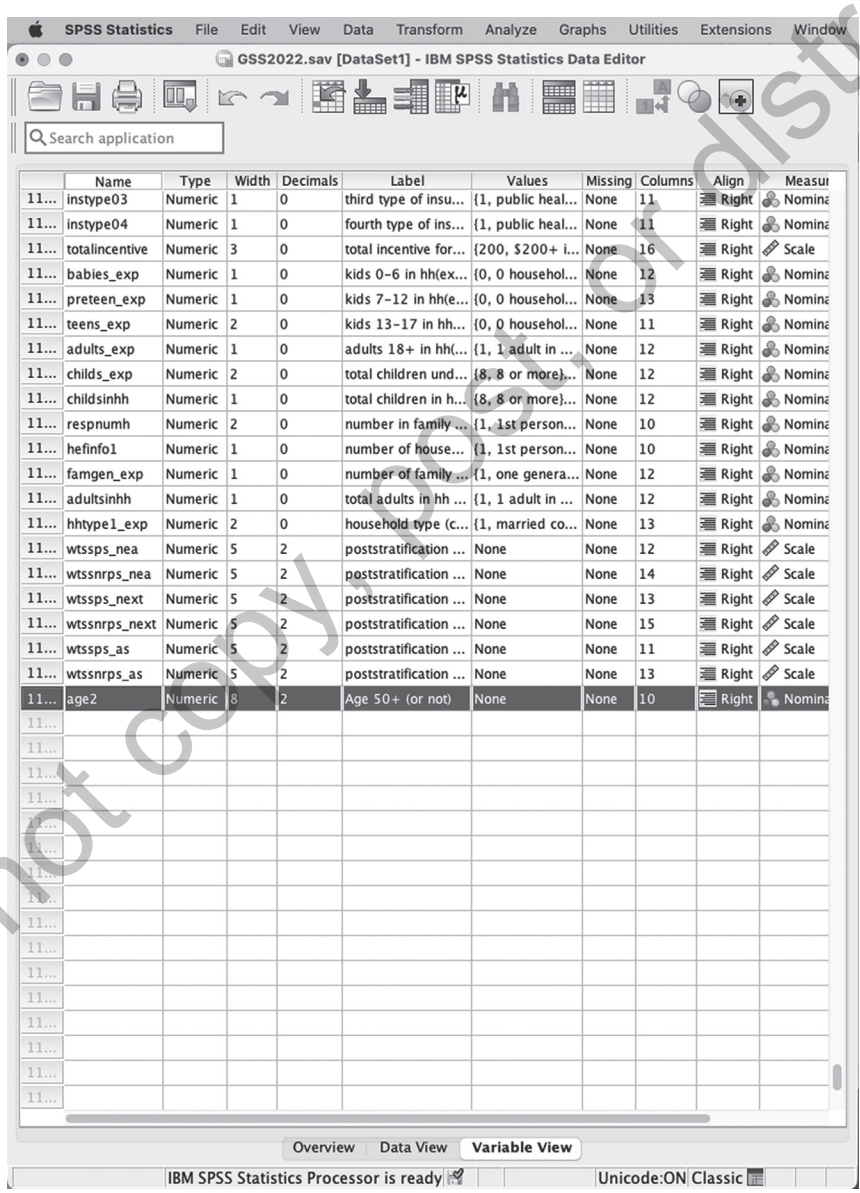
Also, under "Old Value," select "System- or user-missing." Under "New Value" select "System-missing." This will ensure that missing values continue to be treated as such, even if they had been recorded as numeric values. Click the "Add" button once again to confirm that this instruction is added to the list.

Now your instructions have been entered, and you can click the "Continue" button; this will close the current dialog box and return you to the original "Recode into Different Variables" dialog box. Once there, you must click the "OK" button for SPSS Statistics to process your request to recode and then create the new variable.

If the "OK" button is dimmed and SPSS Statistics will not allow you to click it, then one of the preceding steps must not have been completed. The one most often overlooked

is the requirement to click the "Change" button, which adds the new variable name for the output variable.

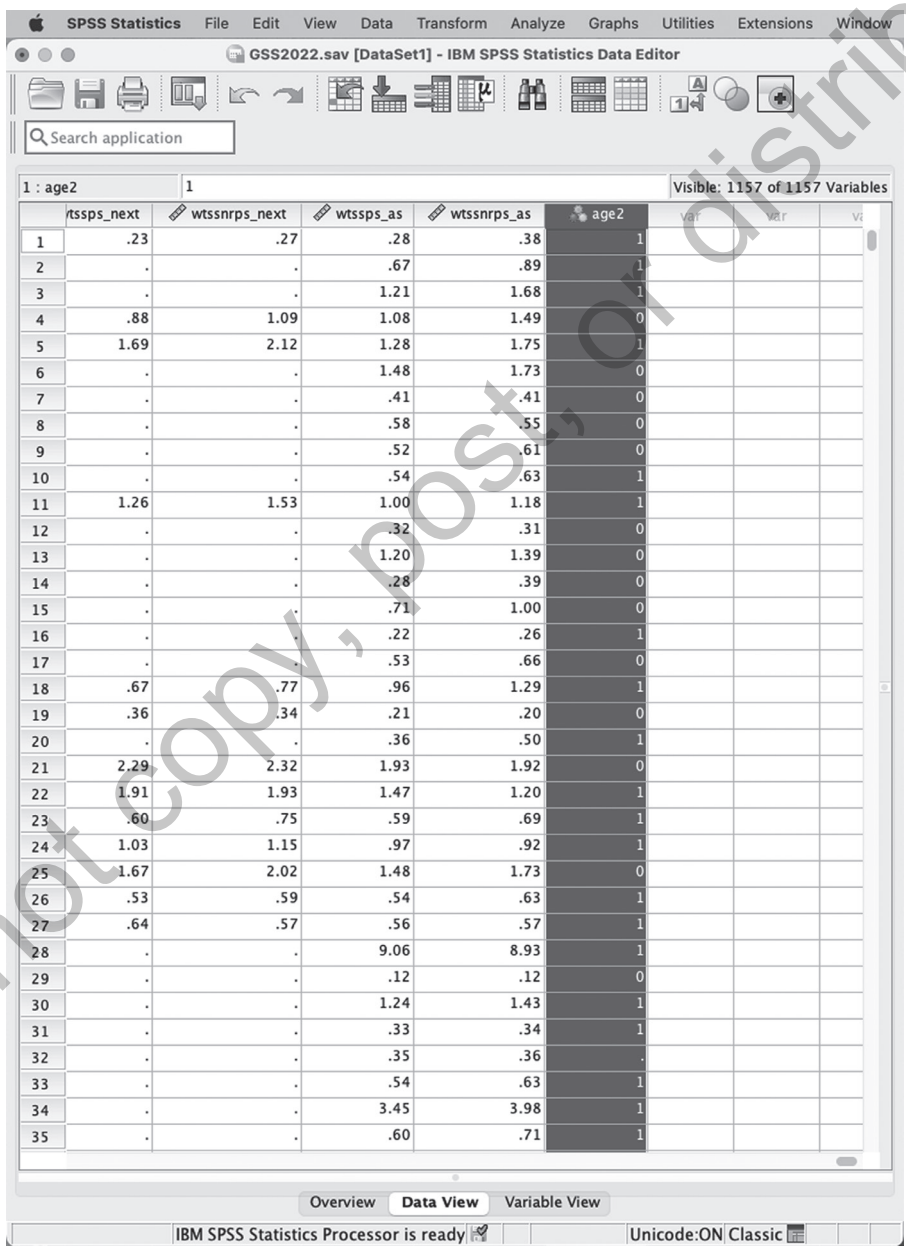
After you have completed the recoding, notice that the new variable is appended to the bottom of the variable list in the Variable View of the Data Editor window. It also happens that the variable is appended to the end (all the way to the right) of the Data View. You can move that variable to another place in the list if you wish by first selecting its entire row and then dragging the row (represented by a red line) up the screen to place it between two of the other variables above. In fact, you can change the "file order" of any of the variables if you wish. This might have utility for you if, for instance, you would like two or more variables to be near each other to compare or verify values in the Data View window.



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measur
11...	instype03	Numeric	1	0	third type of insu...	{1, public heal...	None	11	Right	Nomina
11...	instype04	Numeric	1	0	fourth type of ins...	{1, public heal...	None	11	Right	Nomina
11...	totalincentive	Numeric	3	0	total incentive for...	{200, $200+ i...	None	16	Right	Scale
11...	babies_exp	Numeric	1	0	kids 0-6 in hh(ex...	{0, 0 househol...	None	12	Right	Nomina
11...	preteen_exp	Numeric	1	0	kids 7-12 in hh(e...	{0, 0 househol...	None	13	Right	Nomina
11...	teens_exp	Numeric	2	0	kids 13-17 in hh...	{0, 0 househol...	None	11	Right	Nomina
11...	adults_exp	Numeric	1	0	adults 18+ in hh(...	{1, 1 adult in ...	None	12	Right	Nomina
11...	childs_exp	Numeric	2	0	total children und...	{8, 8 or more}...	None	12	Right	Nomina
11...	childsinhh	Numeric	1	0	total children in h...	{8, 8 or more}...	None	12	Right	Nomina
11...	respnumh	Numeric	2	0	number in family ...	{1, 1st person...	None	10	Right	Nomina
11...	hefinfo1	Numeric	1	0	number of house...	{1, 1st person...	None	10	Right	Nomina
11...	famgen_exp	Numeric	1	0	number of family ...	{1, one genera...	None	12	Right	Nomina
11...	adultsinhh	Numeric	1	0	total adults in hh ...	{1, 1 adult in ...	None	12	Right	Nomina
11...	hhtype1_exp	Numeric	2	0	household type (c...	{1, married co...	None	13	Right	Nomina
11...	wtssps_nea	Numeric	5	2	poststratification ...	None	None	12	Right	Scale
11...	wtssnrps_nea	Numeric	5	2	poststratification ...	None	None	14	Right	Scale
11...	wtssps_next	Numeric	5	2	poststratification ...	None	None	13	Right	Scale
11...	wtssnrps_next	Numeric	5	2	poststratification ...	None	None	15	Right	Scale
11...	wtssps_as	Numeric	5	2	poststratification ...	None	None	11	Right	Scale
11...	wtssnrps_as	Numeric	5	2	poststratification ...	None	None	13	Right	Scale
11...	age2	Numeric	8	2	Age 50+ (or not)	None	None	10	Right	Nomina

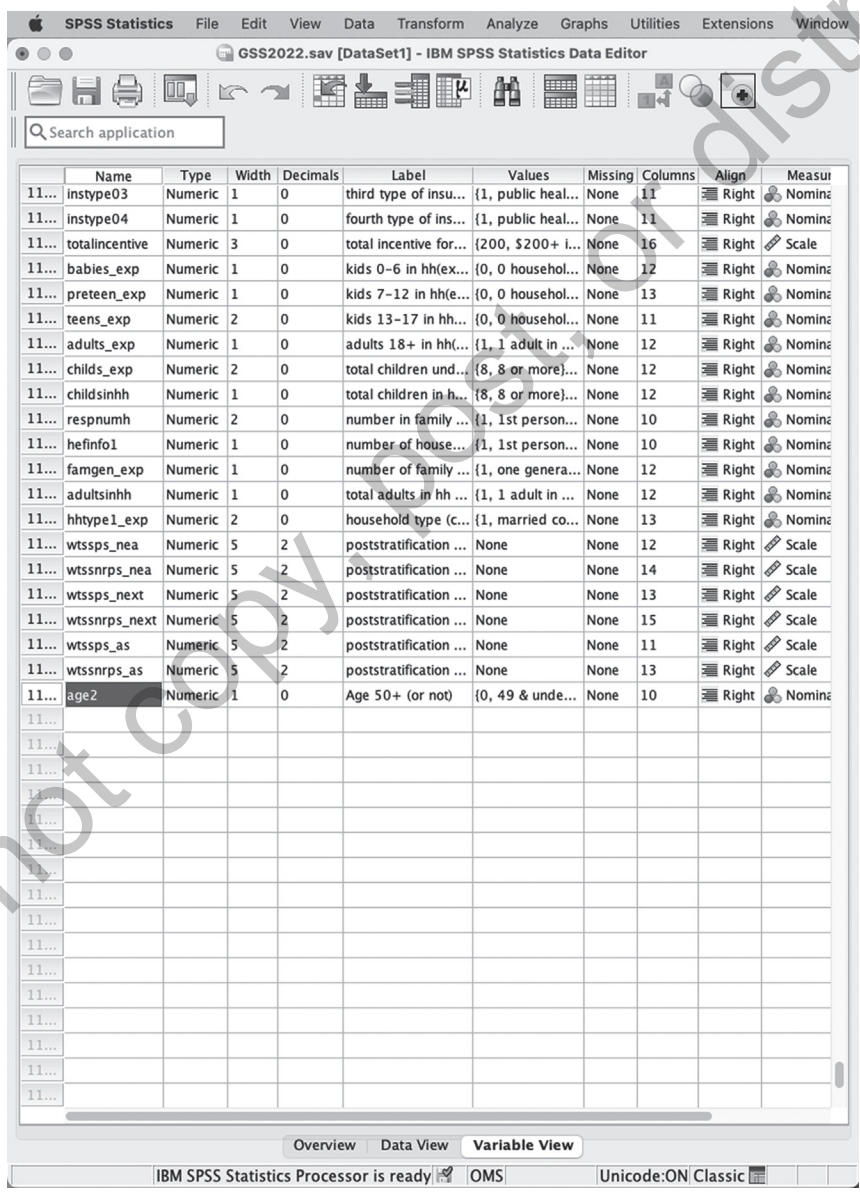
For the variable information, you may wish to change the "Width" column from 8 (the default) to 1 because all possible values are only one digit, including missing case options, such as "IAP" or "Inapplicable" (7), "DK" or "Don't know" (8), and "NA" or "No answer" (9). Similarly, there are no decimals for this variable, so change the "Decimals" column from 2 to 0. Because this variable is dichotomous, select "Nominal" under the "Measure" column.

If you switch to the Data View, you can also see that the variable has been appended at the end of the file (all the way to the right), as shown in the following image:



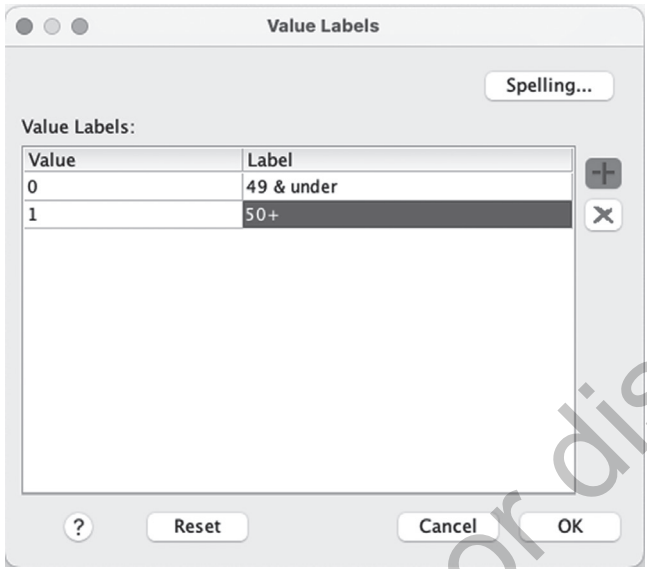
	tssps_next	wtssnrps_next	wtssps_as	wtssnrps_as	age2
1	.23	.27	.28	.38	1
2	.	.	.67	.89	1
3	.	.	1.21	1.68	1
4	.88	1.09	1.08	1.49	0
5	1.69	2.12	1.28	1.75	1
6	.	.	1.48	1.73	0
7	.	.	.41	.41	0
8	.	.	.58	.55	0
9	.	.	.52	.61	0
10	.	.	.54	.63	1
11	1.26	1.53	1.00	1.18	1
12	.	.	.32	.31	0
13	.	.	1.20	1.39	0
14	.	.	.28	.39	0
15	.	.	.71	1.00	0
16	.	.	.22	.26	1
17	.	.	.53	.66	0
18	.67	.77	.96	1.29	1
19	.36	.34	.21	.20	0
20	.	.	.36	.50	1
21	2.29	2.32	1.93	1.92	0
22	1.91	1.93	1.47	1.20	1
23	.60	.75	.59	.69	1
24	1.03	1.15	.97	.92	1
25	1.67	2.02	1.48	1.73	0
26	.53	.59	.54	.63	1
27	.64	.57	.56	.57	1
28	.	.	9.06	8.93	1
29	.	.	.12	.12	0
30	.	.	1.24	1.43	1
31	.	.	.33	.34	1
32	.	.	.35	.36	.
33	.	.	.54	.63	1
34	.	.	3.45	3.98	1
35	.	.	.60	.71	1

Now, because this is a newly created variable, it is very important to insert value labels. If the variable label and/or value labels are not clear, it can be easy to forget what the values are, or in which direction the variable was coded, thereby making the data unusable. In this case, the label for the variable gives a clue if you understand the "(or not)" addendum, which typically means that a one represents the labeled characteristic and that a zero represents the lack of that characteristic; however, let's choose to be safe by eliminating any question for others who may be using this data set at some point in the future. Click the cell in the "Values" column for the variable to which you want to append value labels. Then click the button with three dots, as shown in the following image:



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measu
11...	instype03	Numeric	1	0	third type of insu...	{1, public heal...	None	11	Right	Nomina
11...	instype04	Numeric	1	0	fourth type of ins...	{1, public heal...	None	11	Right	Nomina
11...	totalincentive	Numeric	3	0	total incentive for...	{200, $200+ i...	None	16	Right	Scale
11...	babies_exp	Numeric	1	0	kids 0-6 in hh(ex...	{0, 0 household...	None	12	Right	Nomina
11...	preteen_exp	Numeric	1	0	kids 7-12 in hh(e...	{0, 0 household...	None	13	Right	Nomina
11...	teens_exp	Numeric	2	0	kids 13-17 in hh...	{0, 0 household...	None	11	Right	Nomina
11...	adults_exp	Numeric	1	0	adults 18+ in hh(...	{1, 1 adult in ...	None	12	Right	Nomina
11...	childs_exp	Numeric	2	0	total children und...	{8, 8 or more}...	None	12	Right	Nomina
11...	childsinhh	Numeric	1	0	total children in h...	{8, 8 or more}...	None	12	Right	Nomina
11...	respnumh	Numeric	2	0	number in family ...	{1, 1st person...	None	10	Right	Nomina
11...	hefinfo1	Numeric	1	0	number of house...	{1, 1st person...	None	10	Right	Nomina
11...	famgen_exp	Numeric	1	0	number of family ...	{1, one genera...	None	12	Right	Nomina
11...	adultsinhh	Numeric	1	0	total adults in hh ...	{1, 1 adult in ...	None	12	Right	Nomina
11...	hhtype1_exp	Numeric	2	0	household type (c...	{1, married co...	None	13	Right	Nomina
11...	wtssps_nea	Numeric	5	2	poststratification ...	None	None	12	Right	Scale
11...	wtssnrps_nea	Numeric	5	2	poststratification ...	None	None	14	Right	Scale
11...	wtssps_next	Numeric	5	2	poststratification ...	None	None	13	Right	Scale
11...	wtssnrps_next	Numeric	5	2	poststratification ...	None	None	15	Right	Scale
11...	wtssps_as	Numeric	5	2	poststratification ...	None	None	11	Right	Scale
11...	wtssnrps_as	Numeric	5	2	poststratification ...	None	None	13	Right	Scale
11...	age2	Numeric	1	0	Age 50+ (or not)	{0, 49 & unde...	None	10	Right	Nomina

Then, the "Value Labels" dialog box will appear:



As before with value labels, click the "+" button and then enter both the value and the label for each value and label you enter. When all value and label combinations have been entered, click "OK," and the value labels will be updated. You can inspect the value labels in the Variable View window by clicking the button with three dots on it in the "Values" column of the variable "age2."

RECODING USING TWO OR MORE VARIABLES TO CREATE A NEW VARIABLE

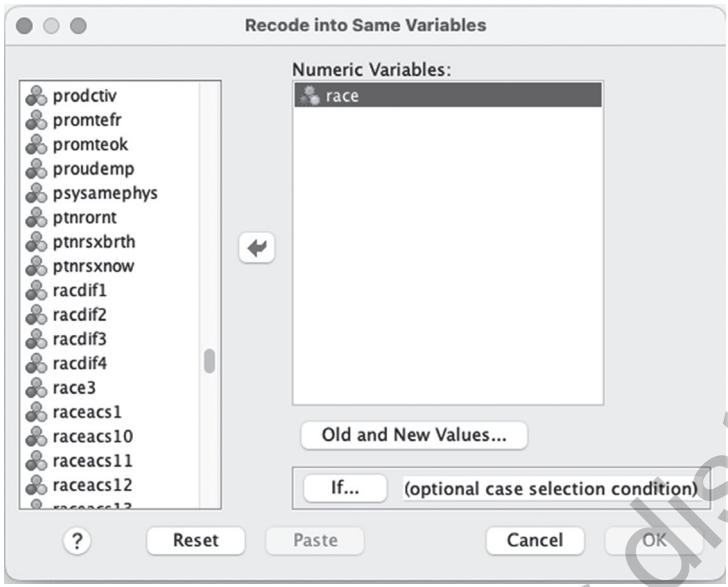
Suppose you want to create a new variable for race/ethnicity that includes information from the variable "race," as well as from the variable "Hispanic." The variable "race" includes only three categories: "White," "Black," and "Other." We want to select those respondents who identify as Hispanic and create a fourth category labeled as such.

In this example, we will recode the race variable into itself using information from the variable "Hispanic." If you would like to preserve the original variable "race," you can do that by duplicating it first; you might call it "race3" because it consists of three categories.

One easy way to do this is to highlight the row (in Variable View) with the variable, "race", then COPY and PASTE the race variable data into the first open row at the end of the file. This sets up the variable and duplicates the labels/categories. Now, COPY the column from the original "race" variable in the Data View and paste it into the already [empty] row labeled "race". Somewhere along the way, you'll need to change the name for one of these two identical race variables to "race3".

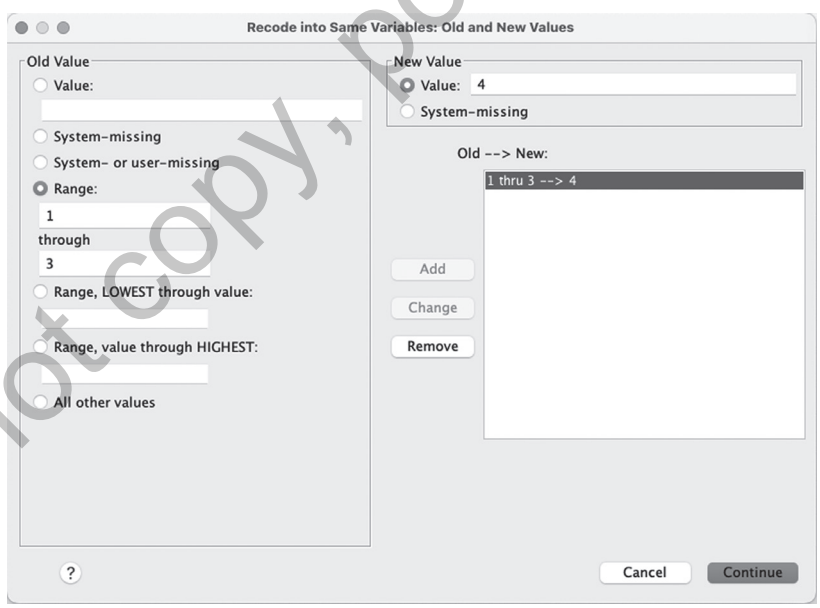
Now, let's begin recoding. Select the following menus in SPSS Statistics:

Transform → Recode into Same Variables . . .



When the "Recode into Same Variables" dialog box appears, select and move the "race" variable to the "Numeric Variables" slot.

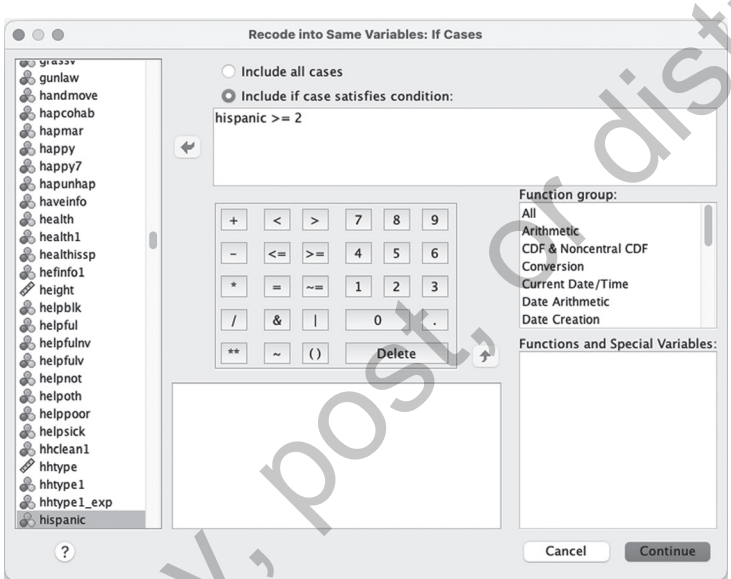
Now, click "Old and New Values . . .". Here, you will make the following entry: Old values from "1" through "3" will be routed to the new value of "4." See the following image to verify your entry. Then click "Continue."

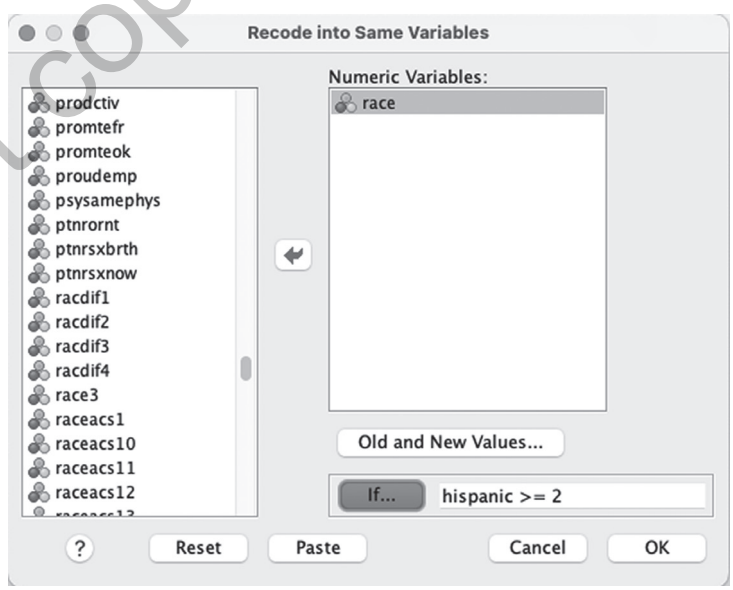


If you think about the logic of what we have just done, we have recoded all respondents (regardless of whether they selected "1," "2," or "3" on the "RACE" variable) to "HISPANIC." But we actually want to recode only the respondents who reported that they were Hispanic. So, let's add that condition to our SPSS Statistics command now.

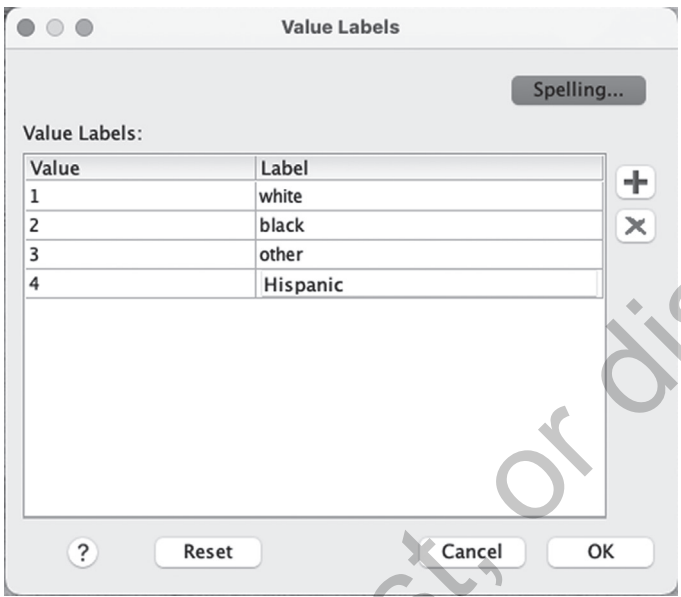
Click the "If . . ." button in the "Recode into Same Variables" dialog box, and you will be presented with the dialog box you see in the first screen image below.

First, select the radio button associated with "Include if case satisfies condition." Now, find the variable "HISPANIC" in the list on the left, and click the right-pointing arrow to send it to the slot under "Include if case satisfies condition." Because the variable "HISPANIC" is coded "1" for not Hispanic, and the values "2" through "50" indicate different types of Hispanic recognition, we want to include those cases in which respondents selected the number "2" or higher. So, click ">=" and then "2." The equation is complete, and you should click "Continue." You will then be taken back to the original dialog box, which will be updated to reflect the new condition that was just entered. Make sure you see that, as pictured in the second screen image below, and then click "OK."



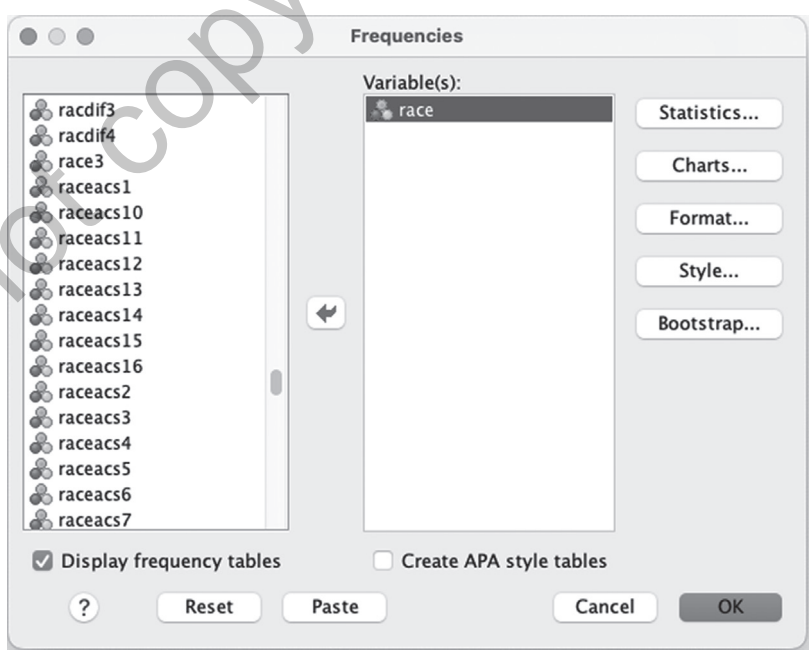


Because we have added a new category, it will be necessary to add the additional value label, as illustrated in the following dialog box. Remember that we get this box by going to the Variable View tab of the Data Editor window, finding the cell in the "Value Labels" column for the variable "RACE," and clicking the button with the three dots.



Click "OK" to add the new value label. Now, to verify that the recoding was done correctly, request a frequency distribution of the new race variable:

Analyze → Descriptive Statistics → Frequencies . . .



You can also ask for frequency distributions of the original race variable, now called "race3" or whatever you decided to rename it, along with the variable "Hispanic" to verify that the recoding was done properly. The distribution described in Table 2.1 reveals that the recoding was carried out as we intended.

TABLE 2.1 ■ Race of Respondent

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	white	2234	53.8	54.7	54.7
	black	715	17.2	17.5	72.2
	other	392	9.4	9.6	81.8
	Hispanic	744	17.9	18.2	100.0
	Total	4085	98.5	100.0	
Missing	System	64	1.5		
Total		4149	100.0		

COMPUTING VARIABLES

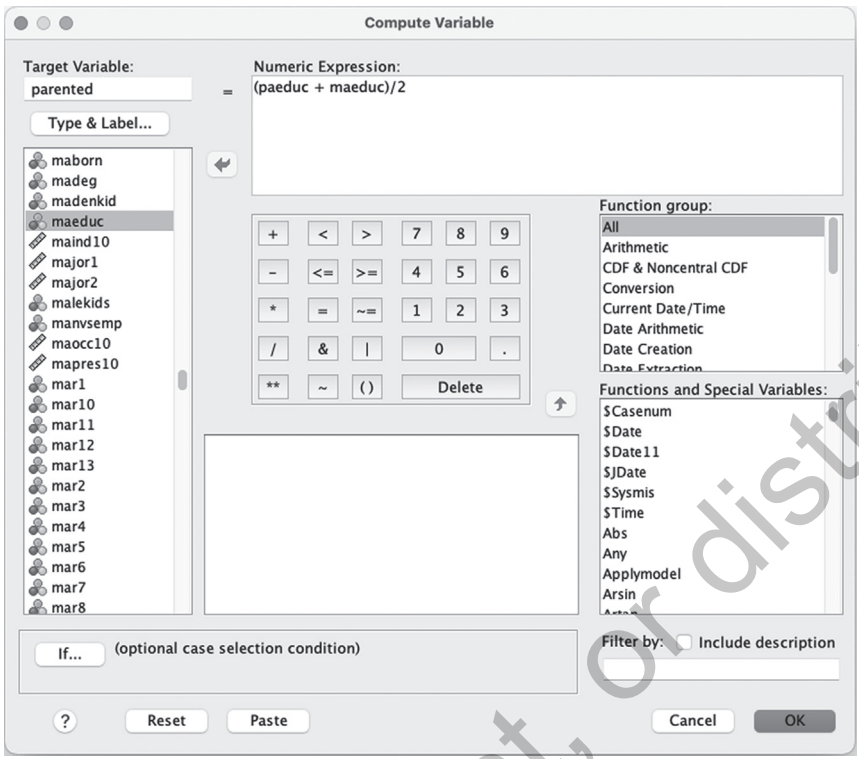
There are numerous reasons why a user of SPSS Statistics would be interested in computing a new variable. For example, one may want to construct an index from individual questions, or one may wish to compute the logarithmic (log) function of a particular variable. In this example, we want to compute the average education level of the respondents' parents. So, we will add the value for mother's education level to that for father's education level and then divide by 2. (Note that in more sophisticated approaches, we might divide by the number of parental responses and disregard sex of parent.)

To perform the computation, select these menus:

Transform → Compute Variable . . .

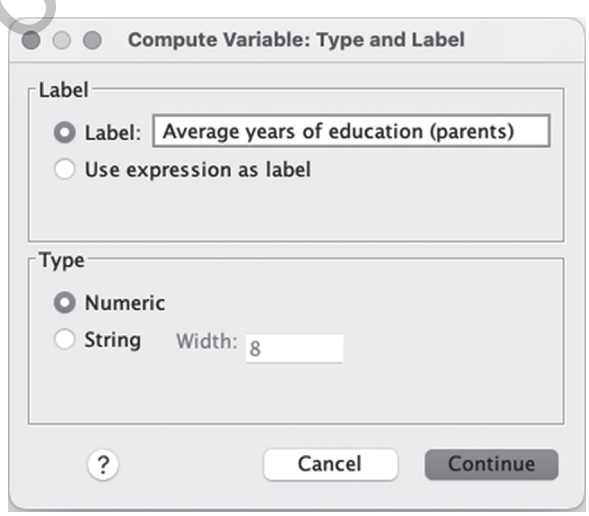
The "Compute Variable" dialog box, shown in the first screen image below, will be presented.

Type the name you wish to assign to the new variable in the "Target Variable" box; in this example, "parented" was chosen. Next, prepare the computation equation in the "Numeric Expression" box. In this case, it is necessary to select the parentheses button "()" first. Then insert the "maeduc" variable by clicking the arrow to the right of the list of variables. Then insert an addition sign by clicking the appropriate button beneath the "Numeric Expression" box. Now, add the "paeduc" variable into the parentheses as well. Next, insert a divisor bar after the parentheses by clicking the button for such, and then click the number "2." This has the effect of adding together the total years of education of both parents and then dividing by 2, yielding the average. After clicking "OK," you will be able to find the new variable, as seen in the third screen image below in the Data View window.

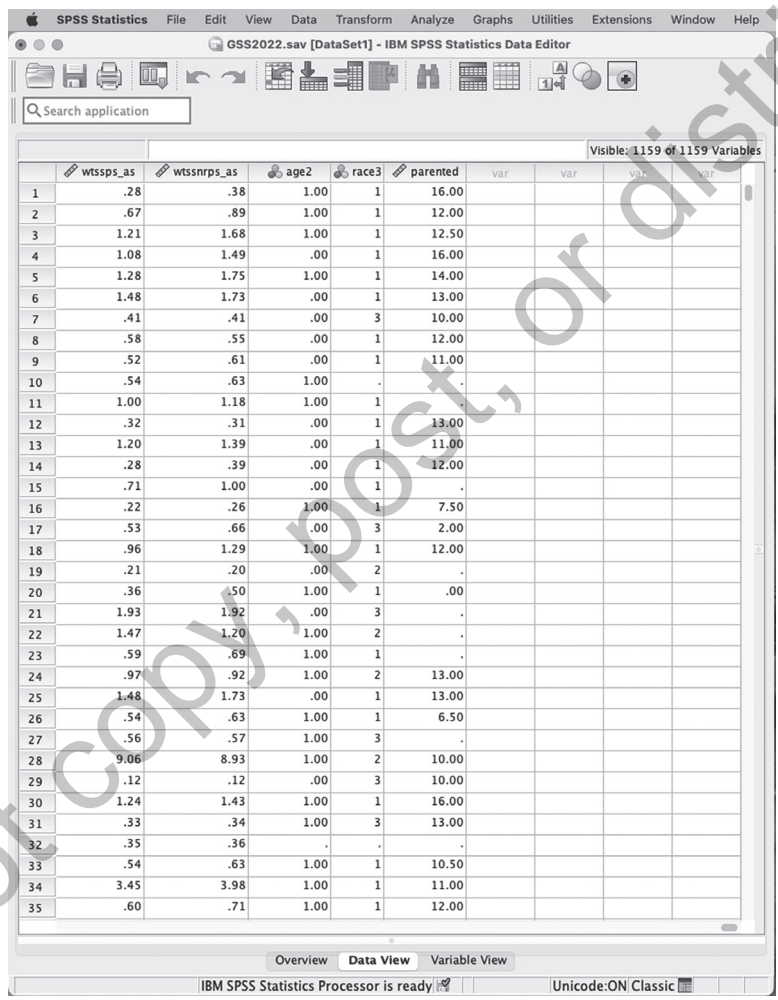


For users who are more comfortable with this process, note that you can type the expression you wish SPSS Statistics to calculate directly into the "Numeric Expression" box. In this case, you could have typed the expression "(maeduc + paeduc) / 2" using your computer keyboard. If you use this method, be very careful to get the spelling of the variable names exactly correct, or SPSS Statistics will not be able to execute your command.

The "Type & Label . . ." button opens a small dialog box that offers an immediate opportunity to enter a variable label and also to define the type of variable (e.g., numeric).



Note the functions that are available: Each selection in the "Function group" pane will bring up a separate list of functions in the "Functions and Special Variables" pane in the lower right corner of the "Compute Variable" box. There are statistical, trigonometric, date, time, string, and other functions that can be used to compute just about anything. Also, if you wish to set up a conditional computation—such that a computation is made in only one case, or there are to be different computations for different cases on the basis of some predetermined condition—then select the "If . . ." button in the lower left corner and enter one or more conditions. The same functions and keypads are provided to instruct SPSS Statistics how to determine the criteria for the conditional computation.

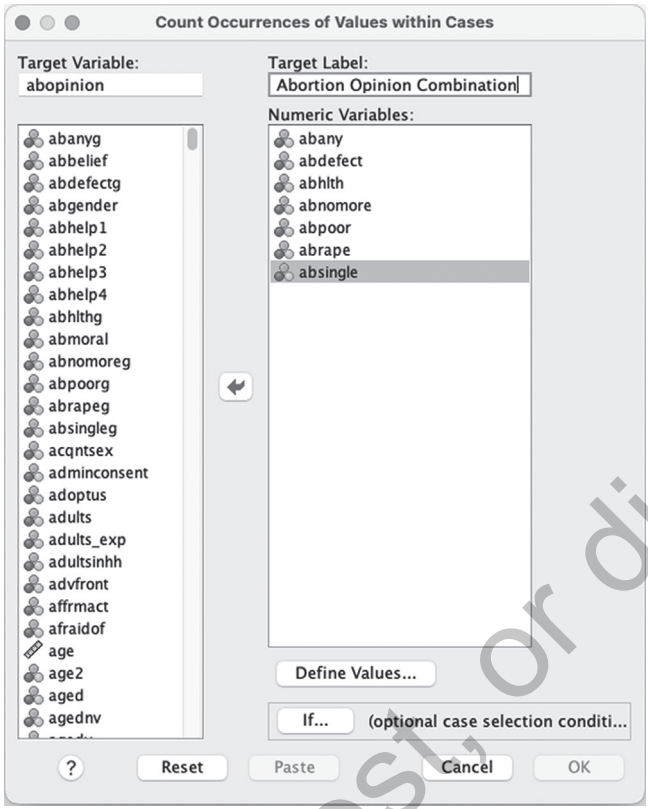


	wtssps_as	wtssnrps_as	age2	race3	parented	var	var	var	var
1	.28	.38	1.00	1	16.00				
2	.67	.89	1.00	1	12.00				
3	1.21	1.68	1.00	1	12.50				
4	1.08	1.49	.00	1	16.00				
5	1.28	1.75	1.00	1	14.00				
6	1.48	1.73	.00	1	13.00				
7	.41	.41	.00	3	10.00				
8	.58	.55	.00	1	12.00				
9	.52	.61	.00	1	11.00				
10	.54	.63	1.00	.	.				
11	1.00	1.18	1.00	1	.				
12	.32	.31	.00	1	13.00				
13	1.20	1.39	.00	1	11.00				
14	.28	.39	.00	1	12.00				
15	.71	1.00	.00	1	.				
16	.22	.26	1.00	1	7.50				
17	.53	.66	.00	3	2.00				
18	.96	1.29	1.00	1	12.00				
19	.21	.20	.00	2	.				
20	.36	.50	1.00	1	.00				
21	1.93	1.92	.00	3	.				
22	1.47	1.20	1.00	2	.				
23	.59	.69	1.00	1	.				
24	.97	.92	1.00	2	13.00				
25	1.48	1.73	.00	1	13.00				
26	.54	.63	1.00	1	6.50				
27	.56	.57	1.00	3	.				
28	9.06	8.93	1.00	2	10.00				
29	.12	.12	.00	3	10.00				
30	1.24	1.43	1.00	1	16.00				
31	.33	.34	1.00	3	13.00				
32	.35	.36	.	.	.				
33	.54	.63	1.00	1	10.50				
34	3.45	3.98	1.00	1	11.00				
35	.60	.71	1.00	1	12.00				

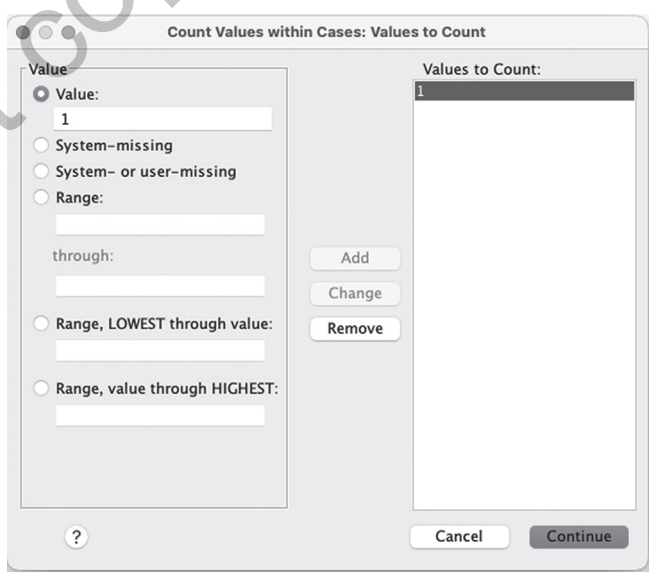
USING THE COUNT FUNCTION

SPSS Statistics allows users the option to add up particular values across different variables. Suppose a researcher wanted to count up the number of instances in which a respondent gave a "yes" answer to particular questions. For this example, consider the GSS (2022) series of questions on opinions relating to abortion. Use the menus below to carry out this example:

Transform → Count Values within Cases . . .

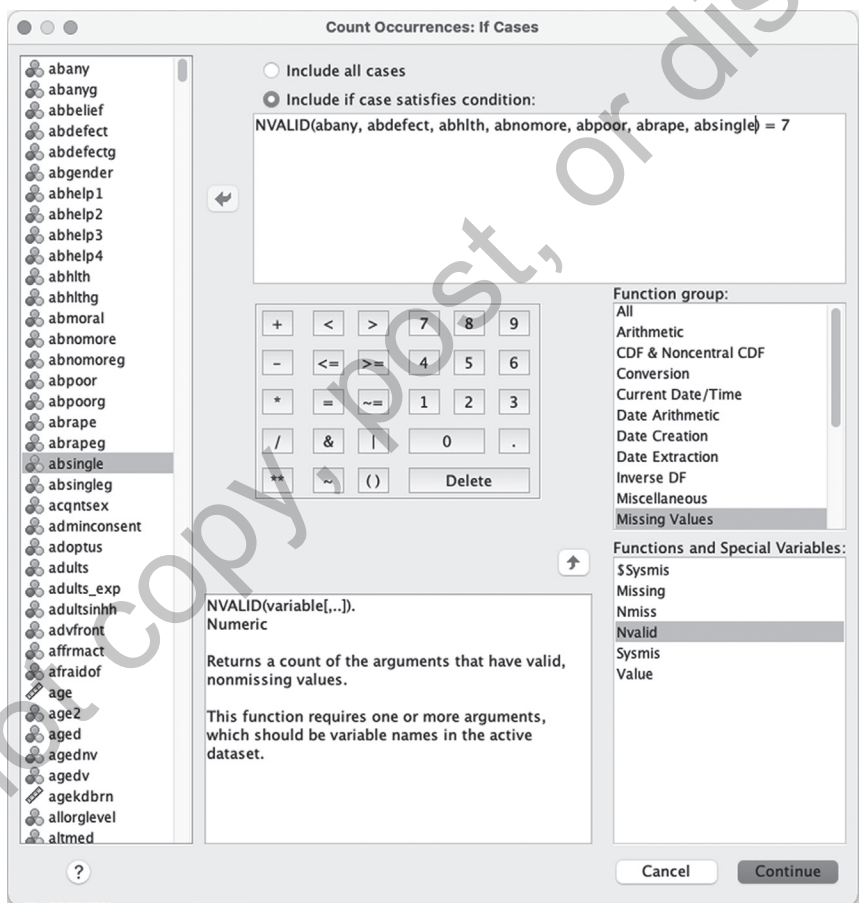


In the dialog box that appears, move the appropriate variables from the variable list on the left into the "Numeric Variables" box. It will also be necessary to enter a name for the new variable to be created in the "Target Variable" slot. The target label can be conveniently entered in the appropriate slot as well. Next, click the "Define Values . . ." button. The following dialog box will be provided:



Here, you will want to select the values to be counted. For the opinion questions that have been selected in this example, a code of "1" indicates an affirmative response (and a "2" indicates a negative response). Therefore, you want to count the number of responses that have a value of "1." Click the radio button next to "Value" at the upper left of the dialog box, and then enter a "1" into the associated slot beneath. Now, click the "Add" button in the middle of the dialog box. The "1" should appear in the "Values to Count" area. Now, click the "Continue" button in this box, and you will be taken back to the "Count Occurrences of Values within Cases" dialog box.

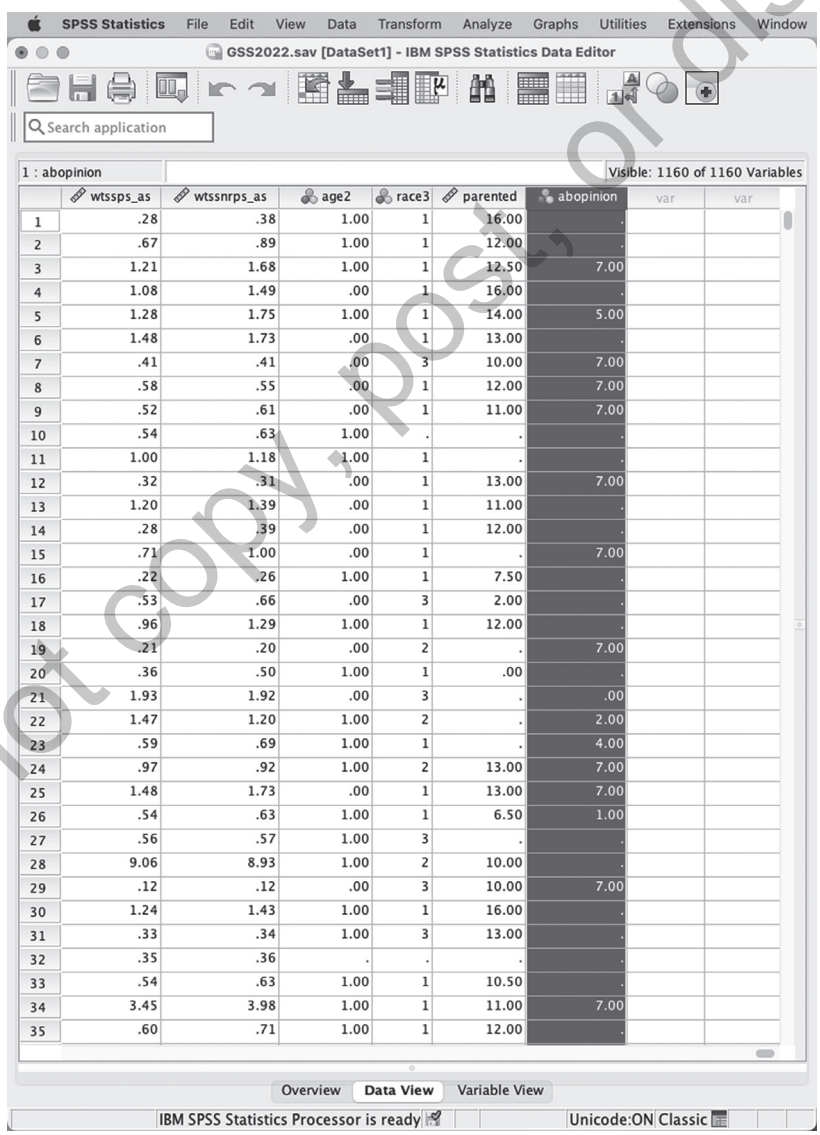
We need to make sure that someone who didn't answer one or more of these seven questions—or who didn't answer any of them—is not counted as someone who "disapproves." That would corrupt our data, and the results from this analysis would not be valid. We can prevent this by using a conditional filter: Click the "If . . ." button, and the following dialog box will appear:



First, select the radio button at the top of the box that corresponds to "Include if case satisfies condition." Next we will prepare the condition. Select "Missing Values" from the "Function group" on the right side of the box. Now, choose "Nvalid" from the list of "Functions and Special Variables." Click the arrow to send it to the box under "Include if case satisfies condition." Then, in the column on the left, find all seven variables having to do with opinions about abortion (abany, abdefect, abhlth, abnomore, abpoor, abrape,

and absingle) and drag them between the parentheses next to "Nvalid," each separated by a comma. This will return a count for each variable that has a valid response. Therefore, if the value returned is a "7," that means that all seven variables contained valid responses for that case. We want all seven items to have valid responses; otherwise, our new variable might include truncated values (e.g., if someone answered only three questions and approved on all three, the new variable would make it look as if he or she approved of only three out of seven, or fewer than half, when the respondent actually approved in all cases for which he or she replied). So, set this expression equal to "7," as displayed in the previous illustration.

Now click "Continue" and then the "OK" button in the prior dialog box, to which you will be returned after this one closes. Other than a record of the SPSS commands executed, no output will be generated. A new variable, however, will be created. See the following screen image for the data contained within the new variable that has been created:



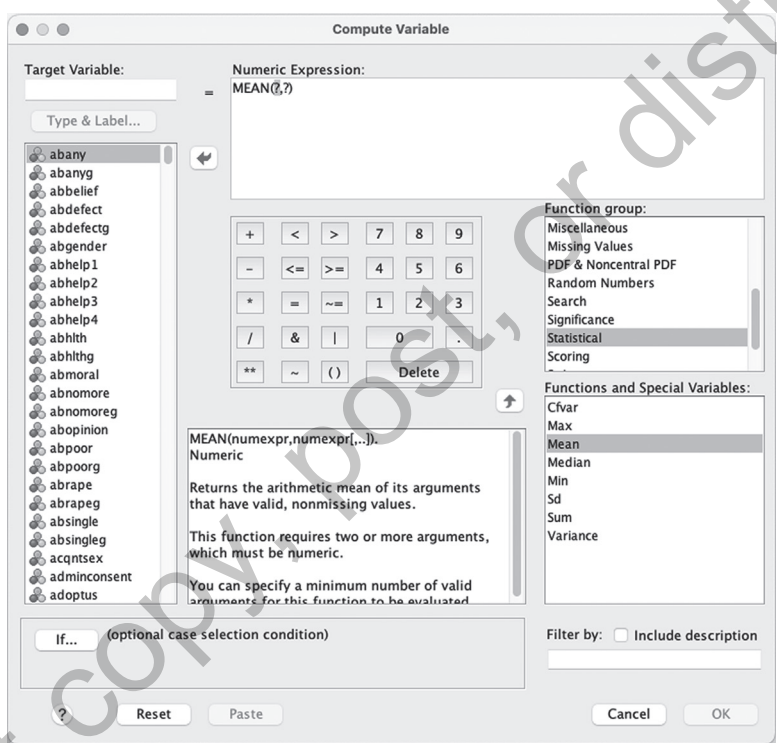
	wtssps_as	wtssnrps_as	age2	race3	parented	abopinion	var	var
1	.28	.38	1.00	1	16.00	.		
2	.67	.89	1.00	1	12.00	.		
3	1.21	1.68	1.00	1	12.50	7.00		
4	1.08	1.49	.00	1	16.00	.		
5	1.28	1.75	1.00	1	14.00	5.00		
6	1.48	1.73	.00	1	13.00	.		
7	.41	.41	.00	3	10.00	7.00		
8	.58	.55	.00	1	12.00	7.00		
9	.52	.61	.00	1	11.00	7.00		
10	.54	.63	1.00	.	.	.		
11	1.00	1.18	1.00	1	.	.		
12	.32	.31	.00	1	13.00	7.00		
13	1.20	1.39	.00	1	11.00	.		
14	.28	.39	.00	1	12.00	.		
15	.71	1.00	.00	1	.	7.00		
16	.22	.26	1.00	1	7.50	.		
17	.53	.66	.00	3	2.00	.		
18	.96	1.29	1.00	1	12.00	.		
19	.21	.20	.00	2	.	7.00		
20	.36	.50	1.00	1	.00	.		
21	1.93	1.92	.00	3	.	.00		
22	1.47	1.20	1.00	2	.	2.00		
23	.59	.69	1.00	1	.	4.00		
24	.97	.92	1.00	2	13.00	7.00		
25	1.48	1.73	.00	1	13.00	7.00		
26	.54	.63	1.00	1	6.50	1.00		
27	.56	.57	1.00	3	.	.		
28	9.06	8.93	1.00	2	10.00	.		
29	.12	.12	.00	3	10.00	7.00		
30	1.24	1.43	1.00	1	16.00	.		
31	.33	.34	1.00	3	13.00	.		
32	.35	.36	.	.	.	.		
33	.54	.63	1.00	1	10.50	.		
34	3.45	3.98	1.00	1	11.00	7.00		
35	.60	.71	1.00	1	12.00	.		

Note that the values, except of course for missing cases, are all appended with ".00." You will want to go back to the Variable View tab of the Data Editor screen to clean up the details, such as width, decimals (change from 2 to 0), and measure.

COMPUTING AN INDEX USING THE MEAN

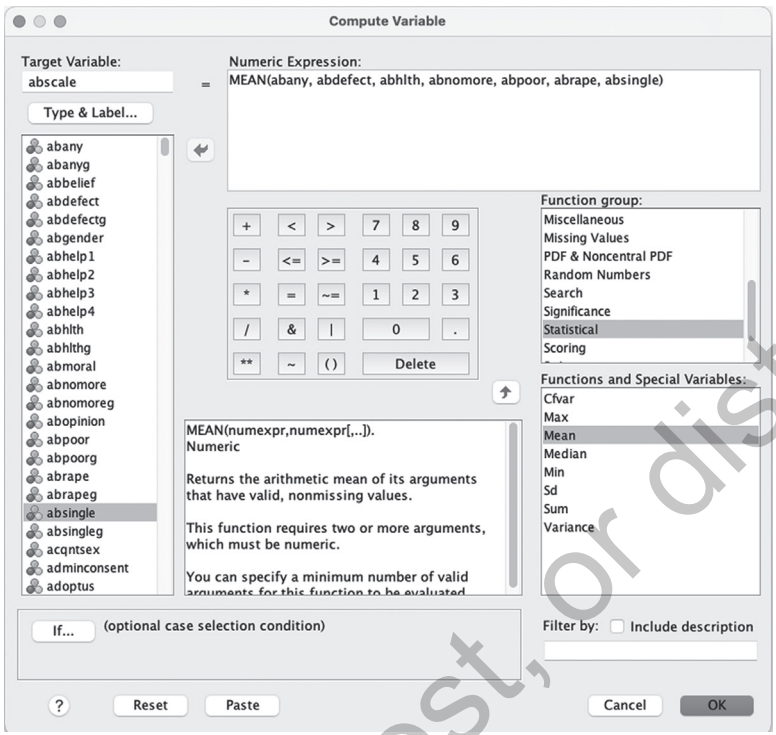
An index can be constructed using the "Compute" command in SPSS Statistics. The most direct way of doing this is to use the "Mean" function. To use this method, click these menus:

Transform → Compute Variable . . .



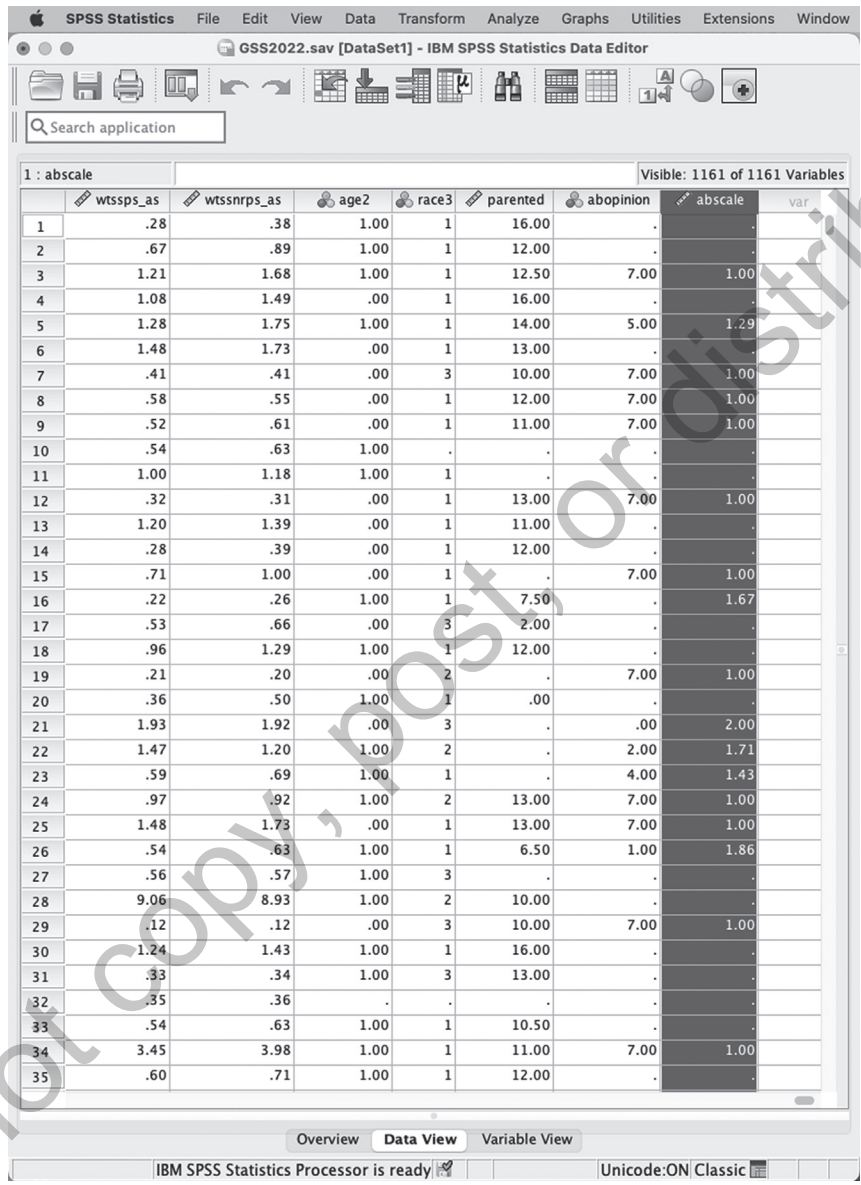
When presented with this dialog box, first note that if any variables, functions, and so on remain from prior use of the Compute Index tool, be sure to click the "Reset" button. Now, in the "Compute Variable" dialog box, select "Mean" from the "Functions and Special Variables" list in the lower right corner. If you don't see "Mean," make sure the "Function group" list above is set to either "Statistical" or "All." It's easier to find within the "Statistical" list, but choosing "All" guarantees that all functions will be found within the very long list provided. After you select "Mean," click the up arrow, which will send "Mean" to the pane under "Numeric Expression." It will appear as it does in the preceding image. You must then insert all the variables of interest within the parentheses after "MEAN" (replacing the question

marks), each separated by a comma. See the following dialog box for how this is done in the current example:



After selecting the variables from the bank on the left and dragging them into the parentheses, enter the name of the target variable (the new variable to be created, in this case, "abscale") in the pane under "Target Variable." Then click the "Type & Label . . ." button. The following short dialog box will be provided:

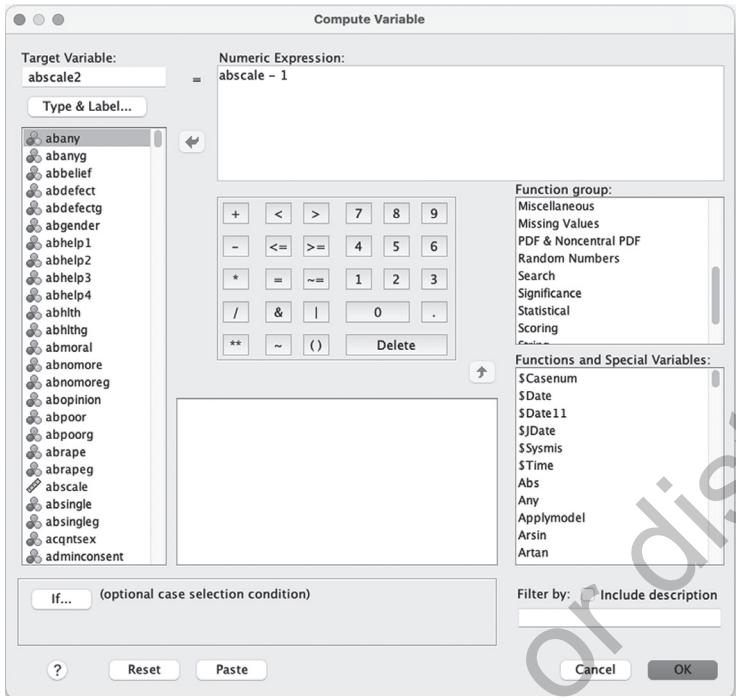
Click the radio button next to "Label" and enter the appropriate variable label in the adjacent slot. Now, click "Continue" and then "OK." A new variable, "abscale," will be created and appear in the SPSS Statistics Data Editor window as shown below:



	wtssps_as	wtssnrps_as	age2	race3	parented	abopinion	abscale	var
1	.28	.38	1.00	1	16.00	.	.	
2	.67	.89	1.00	1	12.00	.	.	
3	1.21	1.68	1.00	1	12.50	7.00	1.00	
4	1.08	1.49	.00	1	16.00	.	.	
5	1.28	1.75	1.00	1	14.00	5.00	1.29	
6	1.48	1.73	.00	1	13.00	.	.	
7	.41	.41	.00	3	10.00	7.00	1.00	
8	.58	.55	.00	1	12.00	7.00	1.00	
9	.52	.61	.00	1	11.00	7.00	1.00	
10	.54	.63	1.00	.	.	.	.	
11	1.00	1.18	1.00	1	.	.	.	
12	.32	.31	.00	1	13.00	7.00	1.00	
13	1.20	1.39	.00	1	11.00	.	.	
14	.28	.39	.00	1	12.00	.	.	
15	.71	1.00	.00	1	.	7.00	1.00	
16	.22	.26	1.00	1	7.50	.	1.67	
17	.53	.66	.00	3	2.00	.	.	
18	.96	1.29	1.00	1	12.00	.	.	
19	.21	.20	.00	2	.	7.00	1.00	
20	.36	.50	1.00	1	.00	.	.	
21	1.93	1.92	.00	3	.	.00	2.00	
22	1.47	1.20	1.00	2	.	2.00	1.71	
23	.59	.69	1.00	1	.	4.00	1.43	
24	.97	.92	1.00	2	13.00	7.00	1.00	
25	1.48	1.73	.00	1	13.00	7.00	1.00	
26	.54	.63	1.00	1	6.50	1.00	1.86	
27	.56	.57	1.00	3	.	.	.	
28	9.06	8.93	1.00	2	10.00	.	.	
29	.12	.12	.00	3	10.00	7.00	1.00	
30	1.24	1.43	1.00	1	16.00	.	.	
31	.33	.34	1.00	3	13.00	.	.	
32	.35	.36	.	.	.	.	.	
33	.54	.63	1.00	1	10.50	.	.	
34	3.45	3.98	1.00	1	11.00	7.00	1.00	
35	.60	.71	1.00	1	12.00	.	.	

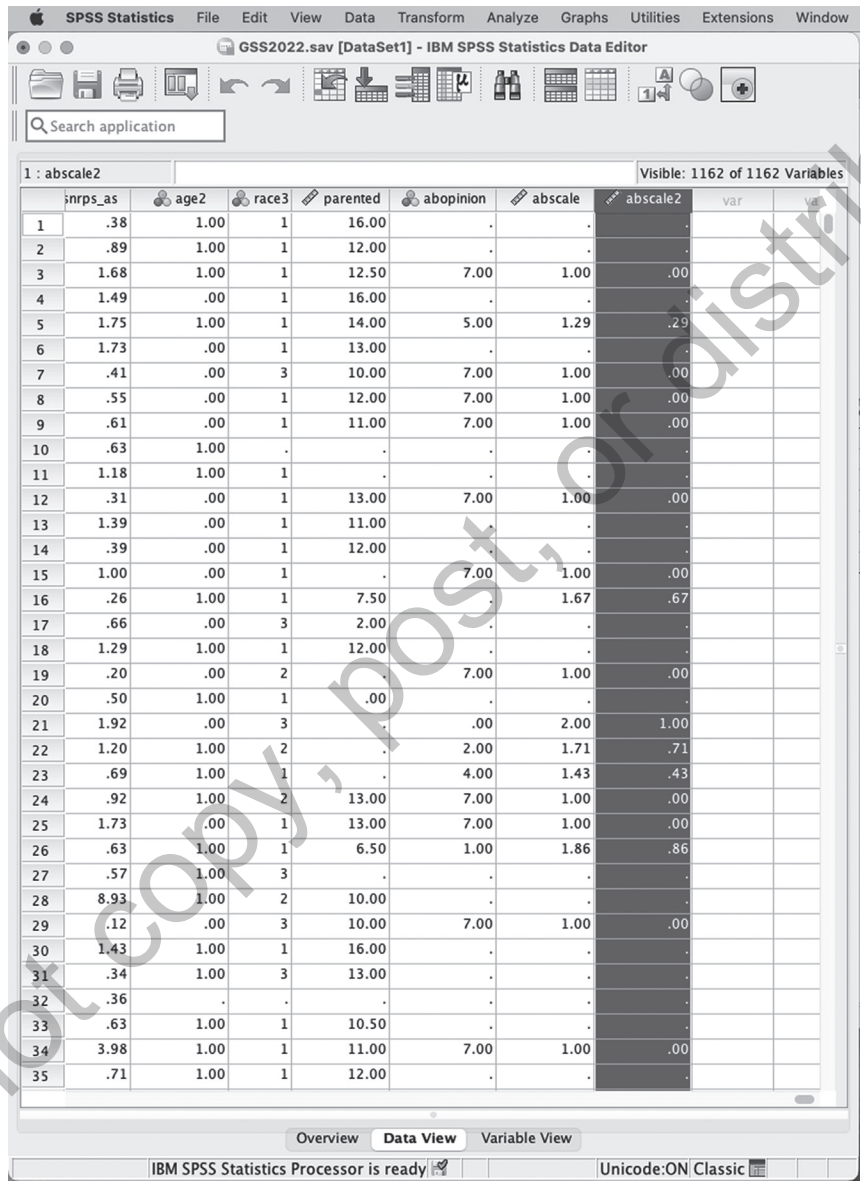
In this case, for nonmissing data (valid cases), notice that the value computed for the mean for each subject falls between (and includes) 1 and 2. This is because 1 represents "yes" and 2 represents "no." For statistical purposes, it is sometimes beneficial to have a result between 0 and 1 instead. To arrange for this, you could add a command to subtract the number 1 from the completed mean function in the original "Compute Variable" box, or you can go back now and make the change, as follows. Again, click these menus:

Transform → Compute Variable . . .



If the dialog box presented is populated with data, click the "Reset" button at the bottom of the box. Now, enter a name for the new variable. (Although it is not recommended, you could overwrite the original variable name.) Click "Type & Label . . .":

Drag the original variable from the bank on the left to the "Numeric Expression" area. Then, on the calculator-style keypad, click "-" and then "1." This will subtract 1 from each case and move the data means into the desired range (between 0 and 1), as they are in the following screen image:

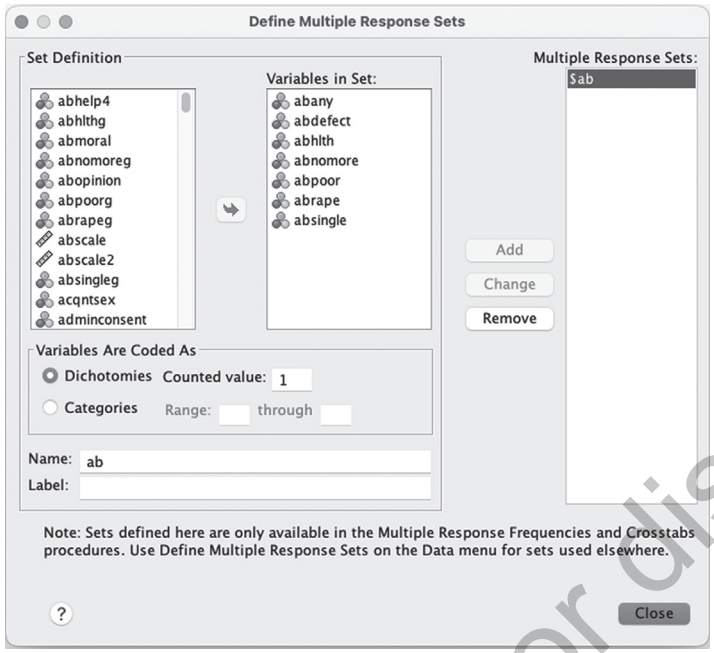


	snrps_as	age2	race3	parented	abopinion	abscale	abscale2	var	var
1	.38	1.00	1	16.00	.	.	.		
2	.89	1.00	1	12.00	.	.	.		
3	1.68	1.00	1	12.50	7.00	1.00	.00		
4	1.49	.00	1	16.00	.	.	.		
5	1.75	1.00	1	14.00	5.00	1.29	.29		
6	1.73	.00	1	13.00	.	.	.		
7	.41	.00	3	10.00	7.00	1.00	.00		
8	.55	.00	1	12.00	7.00	1.00	.00		
9	.61	.00	1	11.00	7.00	1.00	.00		
10	.63	1.00	.	.	.	.	.		
11	1.18	1.00	1	.	.	.	.		
12	.31	.00	1	13.00	7.00	1.00	.00		
13	1.39	.00	1	11.00	.	.	.		
14	.39	.00	1	12.00	.	.	.		
15	1.00	.00	1	.	7.00	1.00	.00		
16	.26	1.00	1	7.50	.	1.67	.67		
17	.66	.00	3	2.00	.	.	.		
18	1.29	1.00	1	12.00	.	.	.		
19	.20	.00	2	.	7.00	1.00	.00		
20	.50	1.00	1	.00	.	.	.		
21	1.92	.00	3	.	.00	2.00	1.00		
22	1.20	1.00	2	.	2.00	1.71	.71		
23	.69	1.00	1	.	4.00	1.43	.43		
24	.92	1.00	2	13.00	7.00	1.00	.00		
25	1.73	.00	1	13.00	7.00	1.00	.00		
26	.63	1.00	1	6.50	1.00	1.86	.86		
27	.57	1.00	3	.	.	.	.		
28	8.93	1.00	2	10.00	.	.	.		
29	.12	.00	3	10.00	7.00	1.00	.00		
30	1.43	1.00	1	16.00	.	.	.		
31	.34	1.00	3	13.00	.	.	.		
32	.36	.	.	.	.	.	.		
33	.63	1.00	1	10.50	.	.	.		
34	3.98	1.00	1	11.00	7.00	1.00	.00		
35	.71	1.00	1	12.00	.	.	.		

MULTIPLE RESPONSE

To produce multiple response values (e.g., frequency values combined across multiple variables), choose the following menus:

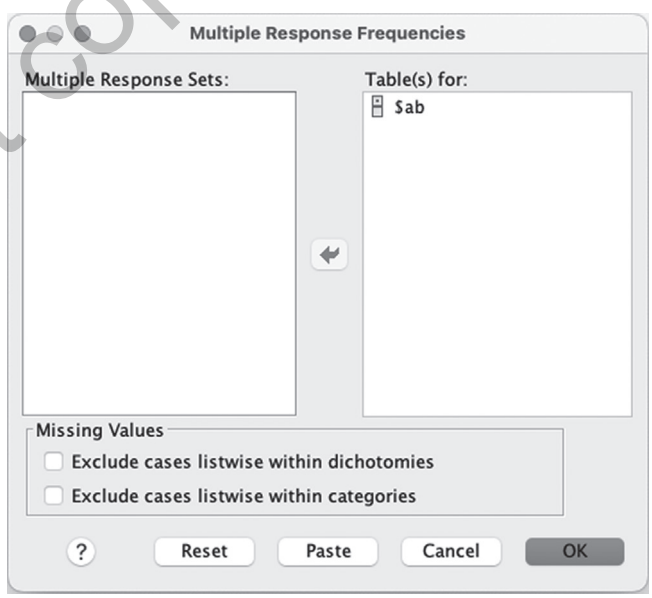
Analyze → Multiple Response → Define Variable Sets . . .



From the variable bank on the left, move all the desired variables into the "Variables in Set" box. In this example, the variables chosen are dichotomies, and we are counting the "yes" value, 1. It is also possible to select a category and a range. Now, name the set and type the name in the "Name" slot. You also have the opportunity to add a label at this time. Finally, click the "Add" button on the right side of the dialog box. You can now click "Close."

To produce a frequency table for the response set that has just been identified, click the following menus:

Analyze → Multiple Response → Frequencies . . .



Select the response set from the list on the left. (In this case, it was the only item in the list.) Move the set to the "Table(s) for" box. Then, click "OK." SPSS Statistics will produce output like the following in Table 2.2 and Table 2.3).

TABLE 2.2 ■ Case Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
$ab[a]	1290	31.1%	2859	68.9%	4149	100.0%

[a]Dichotomy group tabulated at value 1.

TABLE 2.3 ■ $ab Frequencies

		Responses		Percent of Cases
		N	Percent	
$ab[a]	abortion if woman wants for any reason	799	11.9%	61.9%
	strong chance of serious defect	1068	15.9%	82.8%
	woman's health seriously endangered	1253	18.6%	97.1%
	married—wants no more children	819	12.2%	63.5%
	low income—cant afford more children	823	12.2%	63.8%
	pregnant as result of rape	1162	17.3%	90.1%
	not married	799	11.9%	61.9%
Total		6723	100.0%	521.2%

[a]Dichotomy group tabulated at value 1.

Note that the multiple response command allows the easy production of a table combining similar-style variables counting a particular category or range. In the preceding, it is easy to see the similarities and differences in percentages of those who support abortion in the listed circumstances. This frequency table yields the percentage of respondents (here called "Percent of Cases") who approve of abortion in the particular case described. Note, for instance, that 96.1% approve when a woman's health is seriously endangered, whereas only 46.7% approve when the reason is that a woman is not married.